

Lessons of the hundred-hour war

The ground phase of the Gulf War began and ended in less than one turbulent week, leaving a crop of loose ends that need urgently to be sewn together.

THE first lesson is that there is no limit to the destructiveness of conventional weapons, at least in the hands of sophisticated military powers; the highway north from Kuwait to Basra is a pathetic proof of that. But it is also clear that the more conspicuously successful of the conventional weapons used in the Gulf War — cluster bombs, guided bombs and antimissile missiles — are likely for the time being to be useful only to military powers operating such a diversity of military aircraft that the weapons can be used safely. Yet that may be a temporary state of affairs. The war will have stimulated the search for technical developments of two kinds: the range of these weapons, and the ease with which they can be operated from a distance, will be increased, while their inherent sophistication will be further enhanced. The use of laser beams to guide 'smart' bombs to their targets plainly makes their launching platforms dangerously vulnerable, for example; built-in guidance is feasible, preferable and probably near at hand.

So there will now be more military spending on the development of conventional weapons, at least by the substantial military powers (countries such as Britain and France as well as the superpowers). That is bad news for the civil scientific enterprise; if a 'peace dividend' of any kind emerges from rapprochement between the superpowers, less of it than had been hoped will consist of the transfer of research skill from military to civil pursuits. It is also potentially bad news for those who hope that the new weapons will be deployed only within the narrow circle of those who make them. One resounding lesson of last week's war must surely be that military powers embarking on the development of sophisticated conventional weapons must be dissuaded from offsetting the high cost of development by selling the products on the nearly-open market. A formal agreement that the arms trade will now be drastically run down must be in everybody's interests.

Chemicals

Unconventional weapons — chemical and biological weapons particularly — are another matter. Iraq will now be under considerable pressure to abandon the steps it has taken towards these troublesome goals, and will eventually have to comply. Yet Iraq's demonstration that even the threat to use chemical weapons seriously incommoded more sophisticated opponents will not have been lost on others. And while the prospect that chemical weapons might be used is thoroughly odious, it is difficult to see how less sophisticated

powers can equitably be denied the right to manufacture chemical weapons when it may be, or may seem to them, the only way of matching the capability of more technically advanced armies. Chemical weapons may excite particular horror among those threatened with them, but the past few weeks have shown that all ways of killing young men and women are also horrible.

Industrialized governments will probably now prudently agree among themselves not to supply relevant equipment to third parties, but that will not guarantee that weapons of this kind will not spread to determined and technically ingenious countries such as Iraq or Libya. Only a formal and widely accepted agreement can ensure that. What the world needs is not only a treaty such as that which has languished at Geneva for the past several years, but inducements tempting countries with ambitions in these fields to bury them for the sake of the benefits they would derive. By analogy with the Nuclear Non-Proliferation Treaty (NPT), the offer to those who sign of technical assistance with the development of a modern chemical industry would be natural, but might not suffice.

Non-proliferation

The prospect of nuclear proliferation is another issue sharpened by the war. Iraq's near-competence in this field may have been an illusion engendered by its government's habit of wish-fulfilment, but there is a case to be examined (see *Nature* 349, 637; 21 February 1991). In any event, the NPT, the most important of the battery of devices by which the spread of nuclear weapons has been restrained since the beginning of the 1960s, needs to be strengthened in at least two specific ways.

Luckily, the governors of the International Atomic Energy Agency are planning to discuss in June the extension of their safeguards system from nuclear fuel to reactors and other nuclear equipment as well; that, if the signatories of the treaty agree, would be an important step forward. But there remains a need for a mechanism to force crypto-nuclear powers — Israel is merely one of them — to accede to the treaty when there are grounds for believing that their ambiguous status is a regional threat. If, as seems agreed, the authority of the Security Council of the United Nations (UN) has been enhanced by the events of the past few weeks, shall we see it soon exerting pressure to extend the NPT? That would be a logical development.

The lesson that the UN is now a potentially more powerful influence in international relations may be generally ac-



cepted, but will not be easily applied. The majority of the Security Council mustered against Iraq in the past eight months led to the first occasion ever when the UN sanctioned concerted action against the sovereign (if illegal) interests of a member. (The intervention against the North Korean invasion of South Korea in the early 1950s was sanctioned only because the Soviet government was not represented for unrelated reasons.) It would be rash to think that, now, the UN will intervene whenever international injustices crop up. For one thing, as the past few months have shown, the UN is hardly equipped to take up more than one issue at a time. For another, decisions of the Security Council will be swayed by strongly held political inclinations of the permanent members, which have veto rights, until there is enough caselaw to keep prejudice at bay.

But all is not lost. The resolutions against Iraq followed a period of two years during which the two principal superpowers seemed anxious to make the organization function as originally intended — a pacific means of resolving conflict. (Iraq may think itself unlucky not to have annexed Kuwait before this understanding came about.) The UN's usefulness for the years ahead will hang immediately on its members' willingness to ensure, by the array of peace conferences there must be, that the Middle East will become a better place in which to live and work.

Retired politicians have taken to explaining to television audiences why a particular set of interlocking conferences, dealing with, say, military, political and Palestinian matters, might do the trick. Two ingredients are usually missing from these recipes — means of defining and identifying injustice (a 'court' in ordinary language) and an acknowledgement that difficult social and economic problems are most simply solved when they are not defined as zero-sum games (so that each 'winner' does not imply a 'loser'). The UN could do worse than begin by explaining to the Middle East just how prosperous it would be if it could banish the threat of conflict: then the political difficulties would be less formidable. And so, *mutatis mutandis*, elsewhere.

How can such a state of enlightenment be brought about in a region in which conflict seems to be endemic? The best hope is that the experience of the past few weeks will have been cathartic. Already, there are some cheerful signs — the linked presence of Syria in the coalition against Iraq and the more recent renewed confidence of the government of the Lebanon, for example. On the other hand, there is a danger that tension between southern Shiite and northern Sunni Moslems may lead to the partition of Iraq itself, which would hinder long-term security. And it is far from clear how it will be possible simultaneously to engineer the recognition and guarantee of Israel's security and of a land Palestinians can call their own that is not coextensive with Israel itself. But this is not a zero-sum game either. All the parties to a settlement would benefit in some way or another.

Technology could help. It is not so long since, in the 1950s, idealistic Israelis fondly believed that their development of technology, agricultural techniques suited to desert conditions, for example, would win them friends throughout the Middle East.

Several painful decades have shown that the trick would not readily be played, while even attempts to create stronger links between Israel and Egypt have often foundered on the rocks of ancient prejudice. Although all the countries of the region are now better able to look after themselves, it is in no sense unrealistic to hope that well-financed technological developments could do wonders for a desert region with which the whole world is now visually familiar, thanks to the television. Many of the schemes for creating a 'new world order' have an international development bank for the Middle East as an important ingredient. They deserve more support than they are usually given. □

Vitamin quotient

The claim that vitamin supplements increase the IQ of growing children is arresting, but unproven.

WHAT would have happened to Pons and Fleischmann, the inventors of what is called cold fusion, if their press conference on 13 March 1989 had been followed, the same day, by a widely televised account of how the world's need of energy had been met for the rest of time and by the marketing of a handy kit containing a couple of pieces of palladium, a plastic sachet containing electrolyte to be dissolved in water and an explanation of how energy might be won from the contraption by simply connecting it to a supply of electricity? The indulgence, sometimes almost affectionate, with which the scientific community has greeted the delusion of cold fusion would instead have been nasty if its authors had been thought responsible for collecting cash from innocent investors. So what is to be made of last week's bubble of discovery (see page 5) in Britain — the report that cocktails of vitamins formulated in the proportions of the recommended daily doses, but administered to children in modest excess, will increase most children's IQ in a few months or so?

Indulgence

Linus Pauling, one of the cleverest of people, has been saying things like this for decades, and has been dealt with indulgently. Professors emeriti Hans Eysenck (psychology) and John Yudkin (nutrition) have ventured into muddier waters by giving their public blessing to the Californian research said to provide the rationale for the planned sale of vitamin supplements to the general public in Britain. Elsewhere in this issue (page 13) is an appraisal of the article published in the journal *Personality and Individual Differences* (of which Eysenck is the editor); Blinkhorn's review resembles that of a referee whose scepticism of an intended publication is so profound that he has bitten off his tongue for fear of being regarded as incorrigibly intemperate. Yet readers who double as contributors will recognize several "obstacles to publication", as editors tend to call apparently insurmountable flaws.

So is it wrong to claim that extra vitamins increase IQ in growing children? Not necessarily. Given time and opportunity (this is one), the authors might have cogent replies to

Blinkhorn's criticisms. It is also possible that their conclusions are correct, but that their arguments are not.

Moreover, the unfamiliarity of the notion that vitamin dosage might be quickly followed by changes of IQ is not in itself a fatal flaw; to that complaint, the authors could justly retort that the heliocentric hypothesis also at first evoked deep scepticism. But the notion is indeed surprising, and therefore requires extra-meticulous, rather than just plain sloppy, proof.

But is not the conclusion, if correct, important and thus deserving of rapid dissemination? Of its importance (if correct) there can be no doubt; it would revolutionize the education of the young, while educational administrators would quickly tumble to it that vitamin supplements are cheaper than teachers. Yet the price that must be paid by those who make truly great discoveries is the painful wait for recognition. On this occasion, the authors and the foundation of which they are members seem to be seeking a grip on history (and credulous well-intentioned parents) by timing publicity and marketing (by others) to coincide.

It is reprehensible to use what purports to be a serious study in such a way. □

Education in science

Academic scientists keen to help with the school curriculum should look closer to home.

THE cause of education has become to the scientific community what motherhood and apple-pie once were to the United States. And that is not surprising. Governments may be concerned that inadequate public education (not only in science) will undermine economic competitiveness (see *Nature* 349, 2; 3 January 1991), but the research community has an even more practical interest; what will happen to research itself if the steady supply of able students should dry up?

That, no doubt, is the narrow explanation why researchers are increasingly found to have given up time in valuable sabbatical periods to help schools and school systems to develop new science curricula, and why many research laboratories have taken to inviting school students and their teachers for brief internships. (The national laboratories in the United States have a creditable record in this regard.)

The calculation is that young people may thereby be attracted to an absorbing field that they would otherwise overlook. But efforts such as these, laudable though they are, have not spectacularly reversed the downward drift of science enrolments at universities, either in the United States or Europe. Perhaps even more needs to be done. But is it also possible that more radical remedies are required.

Academics are quick to spot the defects of the science curricula at secondary schools — the concern for facts rather than general principles, the preoccupation with humdrum measures of performance and the scant attention paid to laboratory work — but slower to acknowledge the defects of the systems which they themselves administer. The difficulty

is that an education in science consists of two distinct components — a body of knowledge and understanding sufficient to allow a person with imagination to stand, as Newton said he had done, on the shoulders of earlier giants, and an appreciation of what it is to confront problems in the real world whose solutions are unknown.

Most modern science curricula are constructed so that the two elements are somehow blended; what is called project work usually makes an appearance somewhere during an undergraduate curriculum, but often in a somewhat nominal fashion. The digestion of past knowledge tends to constitute the hard core of everything.

Perpetual learning

This imbalance is understandable, even forgivable. The recitation of what is known is easier, while experience shows that all too many undergraduates are perplexed to know how best to come to grips with this material. Yet that is not how mature scientists set about the perpetual learning tasks now unavoidable in research. People faced with the need to use a new technique, experimental or mathematical, do not take several months off in order to train themselves, but rather jump in at the deep end and teach themselves on the job, by an educated process of trial and error. Why should not undergraduates be helped to master their chosen fields in much the same way?

There are several instant objections, of which the simplest is that there are no unsolved problems left that undergraduates can tackle with a reasonable hope of success. But that argument is flawed. For one thing, there is no reason why undergraduates should be indulged, as mature scientists are not, with a guarantee that all questions asked of nature will be neatly solved. And who says that people learn nothing from the frustrations of failure? The other side of that coin is that even novices may often discover truths about the natural world that have somehow eluded others. One wonders, for example, whether the now-fashionable field called chaos would have come to prominence sooner if generations of undergraduates had not been carefully directed towards the soluble problems of linear dynamics. The more serious objection to an education in science more tightly organized around the solution of real problems is that the assessment of the quality of students would then be complicated by the noise engendered by the choice of projects, but that carries less weight than appears. Why should mature scientists be less able to carry out the accurate assessment of their junior colleagues than the assessment of their peers they are always making?

The benefits would be great, and swift. A science curriculum more fully blending study and investigation would give young people a true sense of what science is like and of what it is about. That is bound to be more arresting than the diet on which young people are at present fed. That the same reform would more accurately reflect the characteristic relationship, in the research profession, between practitioners and students is a further benefit. But there would also be more young people willing to throw their lot in with the research community, which is the overall objective. □

US students not taking the Japanese bait

- Student exchange imbalance worsens
- Culture gap remains the problem

Tokyo

THE imbalance in the flow of researchers between Japan and the United States continues to grow, with more than ten times as many Japanese going to the United States as US researchers are coming to Japan, according to new statistics from the Science and Technology Agency (STA). The gap is widening despite Japanese government programmes to encourage Western scientists to visit Japan, and some officials are concerned that the situation could become a renewed source of friction between the two countries.

The unbalanced exchange of researchers became a contentious issue during renegotiation of the US-Japan Science and Technology Agreement in 1987-88, and Japan agreed to take steps to open up its public research laboratories to foreign scientists, in particular those from the United States. Since then, the government has introduced new postdoctoral fellowship schemes, and the number of foreign scientists in Japan is increasing. But the new STA data show that the number of Japanese researchers going to the United States and other Western countries has been growing at a far greater pace.

The STA analysis is based on two sources of data: exit/entry immigration statistics from the justice ministry, and a questionnaire sent to all Japanese national laboratories (except universities and research institutes belonging to the Ministry of Education, Science and Culture).

The justice ministry statistics, which include students and researchers in the social sciences and humanities and scientists on short-term as well as long-term stays, give the widest measure of exchange and show that the number of Japanese going to the United States each year has increased from 18,000 in 1984 to nearly 70,000 in 1989. On the other hand, the reverse flow of researchers from the United States to Japan has increased only from just over 3,000 to 5,252. Similarly, the questionnaire data show that the number of researchers from national laboratories going to the United States increased from 426 to 1,592 during the same period, but the inflow of US researchers to these laboratories increased only from 30 to 119.

The same trend can be seen in data on exchanges with other Western countries. The situation is reversed, however, in the case of South-East Asian countries and China. From those countries, there has been a rapid

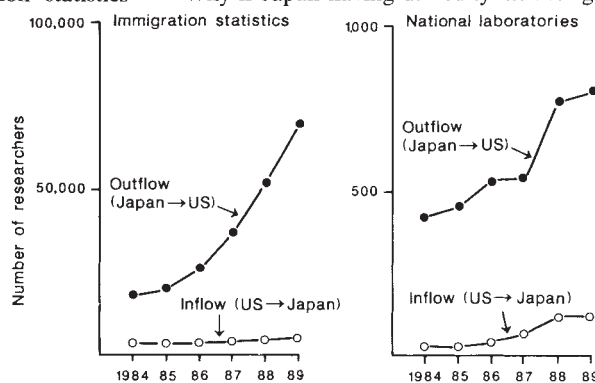
growth in the number of researchers coming to Japan.

STA officials are not sure why there has been such an increase in the number of Japanese researchers going to the United States, but they think it may be due to the rapid increase in value of the yen against the US dollar since 1985. And in the case of researchers at national laboratories, it is probably due in part to a new government fund of ¥200 million (\$1.5 million) for short-term overseas travel introduced in 1987.

To balance the outward exchange, the Japanese government has been actively trying to encourage Western and US scientists to take up the new fellowships it has offered — so much so that some Japanese scientists claim that well qualified South-East Asians, Indians and Eastern Europeans are being rejected in the selection process.

Nevertheless, few US scientists are taking up the awards. When STA introduced a new scheme of 100 fellowships in 1988, they hoped to give half of them to US scientists. But in fiscal year 1990, with the scheme expanded to 160 fellowships per year, only 25 went to US scientists.

Why is Japan having difficulty attracting



Figures from exit/entry immigration statistics (left) tell the same story as an STA questionnaire to national laboratories.

researchers from the United States? Kaname Ikeda of the Ministry of International Trade and Industry (MITI), who helped to set up the STA scheme, says part of the problem is that young US scientists are generally reluctant to go abroad because it will not enhance career prospects at home. Akira Nakanishi, deputy director of STA's international affairs division, echoes this view and adds that there are several other disincentives. Japanese national laboratories, he says, do not provide a particularly attractive environment for Western scientists, visiting researchers have problems with schools for their children, and spouses cannot easily get employment.

David Swinbanks

BSE

First vache folle

Paris

THE first case in France of bovine spongiform encephalopathy (BSE) has been declared on a dairy farm at Plouha, in Brittany. The Ministry of Agriculture and Forests (MAF) says the infected cow comes from British stock and remains the only case so far in the herd.

Following a procedure agreed by MAF last year, the Ministry has bought the entire herd for epidemiological research, at FF2,000 (about \$400) a head over the wholesale abattoir price. It is statistically unlikely, however, that other cases will appear in the same herd, says Dr Brougère, an expert in scrapie at the Maison Alfort veterinary school south of Paris.

The MAF will also try to establish whether the infected cow was given English cattle-feed containing ground offal from ruminant carcasses — a task that Brougère fears will be 'almost impossible'. This feed, thought to be one of the vectors for the agent causing BSE, was banned in Britain from 1988. But it continued to be exported to France until a similar ban was introduced there in 1989.

The MAF does not expect the same kind of epidemic to occur in France as in Britain, as the feed was apparently used in smaller proportions. But Brougère feels it will be only a matter of time before other, 'sporadic' cases appear. With an incubation period of 2-9 years, 'we know there is a risk', she says, 'but we do not know how serious a risk'. One of the problems, she says, has been that veterinary surgeons have not always known the characteristic signs of the illness. But last year a videotape on BSE was sent to all veterinary surgeons, explaining how to distinguish symptoms from those for rabies and hypomagnesia.

Peter Coles

SEAL HUNTING

Moratorium in South Africa

THE South African Minister of Environment Affairs, Louis Pienaar, announced recently that the cabinet had decided to continue the suspension of all commercial seal harvesting on the South African coast until additional research on the issue has been completed and evaluated, which will not be before mid-1992. In doing so, they acted on the recommendation of an independent committee of scientists under the chairmanship of John Hanks, chief executive of the Southern African Nature Foundation.

The committee was appointed to investigate sealing in the wake of the suspension of sealing operations last July, following a public outcry over the granting of a concession to a Taiwanese entrepreneur, Hsien Hsu, to harvest 30,500 Cape fur seals at Kleinsee.

Michael Cherry

IQ data controversy

London

THE authors of a controversial Californian study that claims vitamin and mineral dietary supplements can boost children's intelligence have come under fire from British nutritional scientists, who complain that the scientific community was unable to review the findings before they were publicized on prime-time national television.

The work has also been attacked scientifically for the way its data were analysed and presented (see News and Views, page 13), but even if its conclusions turn out to be correct, the questions surrounding its handling by the authors and the media are likely to remain. In such cases, where a study's findings may affect consumer buying patterns or even government policy, critics say that the researchers should be particularly careful to submit their results to thorough peer review before public release.

The study was led by Stephen Schoenthaler, a California State University criminologist, who has been working on the links between nutrition, behaviour and IQ for a decade. He found a statistically significant increase in IQ among a group of children who were given daily supplements of vitamins and minerals. After 13 weeks, the non-verbal IQ scores of these children had improved by 3.7 points; there was, however, apparently no significant change in verbal IQ scores.

The news of these tantalizing results was out almost before they had appeared in the scientific press. The study was published in the journal *Personality and Individual Differences* (PID) on 26 February, as part of the proceedings of a symposium. The same night, the findings were presented on the British Broadcasting Corporation (BBC) science programme, QED.

Three years ago, QED broadcast a programme that suggested a link between nutrition and IQ, which provoked immense



public concern and scientific controversy in the United Kingdom.

Several aspects of the study's handling are disturbing to other British scientists. Roger Whitehead, director of the Medical Research Council (MRC)'s Dunn Nutrition Unit in Cambridge and president of the

Extraordinary news for parents



In a recent US trial, the average IQ score of children taking a specific multivitamin and mineral formula increased significantly. That formula is now on sale as Healthcrafts Vitachieve. It had apparently helped balance the children's diet allowing them to reach their natural potential. You may have seen the publicity about these tests, which were conducted by the Dietary Research Foundation: an independent body of scientific experts. For written details of the DRF results, call free on 0800 444 858.

Available from Boots



Healthcrafts' newspaper ad.

Nutrition Society, says that when research findings may have profound implications for public health, it is vital that they be properly evaluated before reaching a general audience. "There's a place for this type of television programme, but only in the light of scientific review."

And some researchers contend that the nutrition study was not given a careful enough review. For example, the meeting at which the data were discussed, which was held in San Francisco late last year, was closed to other researchers in the field.

PID, the journal in which the results were published, has a small (less than 1,000) circulation and is edited by Hans Eysenck, a co-author of the study.

Eysenck on the other hand says that the manuscript was subjected to the journal's usual process of peer review. Part of the reason for choosing PID, he says, is that publication had to be timed to precede the BBC broadcast. But Whitehead counters that other journals can publish quickly if findings are valid and important.

Some researchers question the involvement of the BBC with the study. David Southgate, from the Agricultural and Food Research Council's Institute for Food Research in Norwich, and editor of *The British Journal of Nutrition* says he would have liked the manuscript to have been submitted to a peer-reviewed biomedical or nutritional journal, before the programme was made. "I don't really approve of science effectively being published on the television", he says.

The BBC is also marketing a book by

Stanford initiative

Washington

FROM one of the pre-eminent research universities in the United States has come a proposal for sweeping reform of undergraduate education. Last Sunday, 3 March, Donald Kennedy, president of Stanford University, outlined a series of policy changes aimed at improving the teaching of undergraduates and lessening the tension between the research and the teaching duties of faculty members.

It is a common complaint at US universities that tenure and other personnel decisions are made mostly on the basis of a faculty member's academic research, with teaching prowess a poor second. This system excels in turning graduate students into effective research scientists, but critics say that it reduces the quality of education for undergraduates. At Stanford, at least, Kennedy hopes to improve teaching without impairing research.

Kennedy wants to switch the emphasis from quantity to quality in evaluating a scientist's research. His proposal, which must be approved by a faculty advisory board, would limit the number of publications that are taken into account when the university makes decisions on hiring, promotions, or granting tenure.

To encourage better undergraduate teaching, Kennedy both proposed changes in the way faculty members are evaluated and revealed direct financial incentives. Faculty committees should weigh not just research but also scholarly publications such as textbooks and software. And to make the rewards of better teaching palpable, Kennedy announced that \$7 million will be spent on such things as salary increases, \$5,000 awards for outstanding classroom instruction and fellowships for senior faculty members who would like to specialize in teaching. Robert Pool

Schoenthaler, *Improve your Child's IQ and Behaviour*.

Whitehead is also troubled that a vitamin and mineral supplement for children, 'Vitachieve', has been launched in Britain on the basis of the results. That supplement has been endorsed by the Dietary Research Foundation (DRF), the charitable trust that funded the nutrition study. Whitehead is bothered that a supplement is being advertised in Britain on the basis of research done on Californian children. The underlying nutrition of Californian and British children is different, which may affect the results.

Eysenck responds that a study in Cumbria, northwest England, supports the Californian findings, but analysis is not yet complete. Whitehead says the British results should have been published before launching a product on the UK market. "It's absolutely essential that the data collected in Britain will be made available" for nutritionists to draw their own conclusions.

Peter Aldhous

Cetus retains PCR patents

San Francisco

CETUS Corporation has finally nailed down its lucrative patent on the revolutionary polymerase chain reaction (PCR) technology. Last week, a US District Court upheld the validity of the patents granted to the biotechnology company against a challenge by the chemical giant Du Pont.

Although the decision came as no surprise to biotechnology observers, it was nonetheless welcome news to Cetus, whose financial health depends in large part on its rights to the PCR technology. The victory was especially important to the company in the light of the major blow it received last summer, when the US Food and Drug Administration unexpectedly failed to recommend market approval for interleukin-2, its anticancer drug.

The six-week jury trial, which included testimony from three Nobel laureates, examined the origins of PCR, a 'gene amplification' technique. PCR is basically a method of detecting and serially reproducing specific gene fragments from a pool of DNA. It has dramatically transformed many areas of biological research since its introduction in 1985, and has also been used to identify criminals from blood samples, examine archaeological remains and diagnose infections and diseases.

In the trial, Cetus maintained that PCR was invented by its scientist Kary Mullis and was not practised anywhere until Cetus announced it in a 1985 paper. Cetus's patent rights to PCR were first issued in 1987, and were confirmed by a US Patent and Trademark Office re-examination last year.

Du Pont, however, contended that the process was made obvious by the work of Massachusetts Institute of Technology professor and Nobel laureate H. Gobind Khorana and his colleagues in the early 1970s. Two of Khorana's papers and a grant application were presented as evidence by Du Pont, although Khorana declined to testify.

Instead, Du Pont presented as witnesses Stanford University Nobellist Arthur Kornberg and several of Khorana's postdocs from the 1970s, while Cetus countered with testimony from Nobel prizewinners Hamilton Smith of Johns Hopkins University and Sir Aaron Klug of the UK Medical Research Council.

A Du Pont attorney said last week that the company is "reviewing all its options" regarding the decision. A separate phase of the trial, scheduled to begin next week, will consider the Cetus countercharge that Du Pont has infringed upon another PCR-related patent; in that phase, Cetus will seek damages from Du Pont as well as an injunction against Du Pont sales of materials it considers to be PCR technology.

The markets for PCR technology are expected to grow enormously in the next few years. A Cetus joint venture with Perkin-

Elmer Corporation, which sells PCR equipment and reagents to research laboratories, resulted in sales of \$26 million last year. In addition, Cetus has licensed PCR to Hoffmann-La Roche and others to develop diagnostics kits which may be used for a variety of medical purposes including detection of cancers, AIDS and genetic disorders, as well as for forensic identifications. Analysts estimate that the combined PCR market may reach \$400 million within five years.

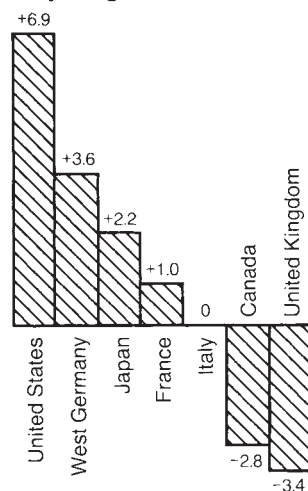
Now that Cetus's rights to that market appear safe, Cetus attorney Peter Staple believes the victory has implications for other biotechnology firms as well. "Patents can be an equalizer", he says. They can allow young, relatively small companies to protect their results from big corporations in the still-unpredictable field of biotechnology.

Elizabeth Schaefer

CITATION ANALYSIS

US science OK

WHILE US science advocates argue that US science is losing ground to the rest of the world, new data from the Institute for Scientific Information (ISI) appear to suggest otherwise. During the past decade, the United States has not only improved nearly 7 per cent in the relative quality of its science (as measured by average citations per US paper compared to the worldwide average), but is also improving faster than any of the other Group of Seven (G-7) nations. The data also show an apparent correlation between funding and science quality. Research budgets in the United States, Japan, Germany and France — whose research has improved relative to the rest of the world — have consistently grown over the decade, while Britain and Canada — whose research quality has slipped over the past five years — have been hit by budget cutbacks.



Research quality of the G-7 nations, expressed as percentage change in citations per paper from 1981-85 to 1986-90, relative to the world average in both years. Source: ISI.

Education under threat

London

THE quality of UK undergraduate education will be damaged if the Universities Funding Council (UFC) continues to insist that universities take on more students, despite the fact that not all of these will be funded, warns the Committee of Vice-Chancellors and Principals (CVCP).

The UFC said last week that it will fund 303,700 undergraduates in 1991-92, some seven per cent more than in 1990-91. But the universities' four-year plans, submitted to the UFC last summer, had requested



Sir Peter Swinnerton-Dyer — grant calculator.

335,000 students in 1991-92. A covering letter from Sir Peter Swinnerton-Dyer, chief executive of UFC, sent to each university to explain how its grant had been calculated, says that the UFC expects universities to admit significant numbers of students without funding from the UFC. Most vice-chancellors regard this as a thinly veiled threat. They say that universities that refuse to expand will be discriminated against in future years when the UFC comes to distribute its budget.

Sir Edward Parkes, chairman of the CVCP, says that this unfunded expansion will mean a decline in the quality of UK undergraduate university education. His comments reflect the official position adopted by the vice-chancellors at their meeting last Friday. But Sir Graham Hills, from Strathclyde University, a long-standing dissident in the vice-chancellors' ranks, says that universities have been claiming for ten years that underfunding will produce an imminent decline in undergraduate education.

Most vice-chancellors accept that the government is unlikely over the next few years to give the UFC sufficient money to fund the desired expansion in full. Last autumn, there was much speculation that some universities would be forced to charge students 'top-up' fees to maintain teaching quality. This would be extremely unpopular with students, and Hills believes the 'political odium' of fees will deter universities from introducing any charges for at least another year. Sir John Kingman, vice-chancellor of the University of Bristol, says the opposition to top-up fees, from the government, as well as from students, means they are unlikely for the 'foreseeable future'.

British university funding has been shrouded in uncertainty since last autumn,

when the UFC's new 'competitive bidding' system for funding undergraduate teaching was suspended (see *Nature* 348, 3; 1 November 1990). Under this system, universities told the UFC how many students they wished to teach by 1994–95, and how much this would cost.

The UFC's logic was that the universities, most of which wish to expand, would compete against each other to take on more students at a reduced cost per student. This would have met the government's twin policy aims of expanding higher education and driving down the average cost of teaching. But university vice-chancellors submitted the vast majority of their bids for more students at the maximum 'guide prices' suggested by the UFC, forcing the UFC to withdraw the system. The UFC says it will unveil a replacement system, which will retain 'an element of competition' between the universities for funds, later this month.

A similar bidding system to that which failed in the universities was introduced for polytechnic funding a year earlier, and the polytechnics have co-operated until now.

INDUSTRIAL INNOVATION

Get academics into the boardroom

London

INCREASED links between academic scientists and industry, including the appointment of university and polytechnic staff to company boards, are recommended today (7 March) as a central plank of the UK House of Lords Science and Technology (S&T) Committee's strategy to revive British industry.

The Lords applaud the efforts that higher education institutes and research council laboratories have taken already to promote technology transfer to industry, although they say that the transfer of technology from government department-funded laboratories has been disappointing. Their report also warns that the government should not use increased academic collaboration with industry as an excuse to reduce further its funding for the UK science base.

The S&T committee's inquiry into innovation in British industry was prompted by the lack of investment in British industry during the 1980s, and the consequent decline in the manufacturing sector. Its report paints a gloomy picture, concluding that British industry is caught within a "vicious circle" of financial disincentives to investment.

The government should take a leading role in recreating a favourable climate for innovation in the United Kingdom, the report says. The government's policy over the past few years of reducing its funding for the development of commercial products and the failure of British public spending on research and development (R&D) to keep pace with that of our industrial competitors — is sending "the wrong signals"

But the Committee of Directors of Polytechnics has now followed the CVCP's lead and declared that further efforts to drive down the cost of teaching each student will undermine teaching quality.

The rejection of the bidding systems across both universities and polytechnics is significant because a merger between their funding bodies is becoming an increasingly likely prospect. The vice-chancellors' argument that further reducing teaching costs will undermine quality would have been hard to sustain in a merged system for higher education funding, if the polytechnics had continued to further improve their efficiency without complaint.

Swinnerton-Dyer told the House of Lords Science and Technology Committee last week that he believes a merger between the UFC and the Polytechnics and Colleges Funding Council (PCFC), at least of their support for teaching, is 'inevitable'. The PCFC is expected this week to approve plans to make its methods of funding more similar to those of the UFC — an important first step in any merger plan.

Peter Aldhous

to industry, the Lords say.

During the past few months, the UK government has implicitly acknowledged these concerns, launching a three-year, £32-million grant scheme to promote R&D in small companies and changing the rules of its LINK research programme, which promotes industrial-academic collaboration in R&D, to encourage more companies to take part. But these changes are likely to have a limited effect, and the S&T committee's report notes that the Department of Trade and Industry's spending on industrial innovation was slashed by more than one half between 1985–86, when £283 million was spent, and 1990–91. The mid-1980s budget should be restored, the report says.

The "short-termism" of industrial companies themselves is similarly berated by the S&T committee. British companies invest too little in R&D, ploughing too much of their profits into dividends for their shareholders, the report says. Investment in R&D is further constrained by the fear that companies reinvesting a large proportion of their profit will be less able to fight off hostile take-over bids.

Investment in R&D could be increased by changing the UK tax laws. At present, companies are allowed to write off the full value of their annual investment in R&D against tax. The S&T committee argues that companies' R&D spending could be increased by making 150 per cent of the increase in a company's R&D spending from one year to the next exempt from tax. Australia introduced a similar measure in 1985, and R&D spending has since increased.

Peter Aldhous

What's in it for science?

London

IN general, British science students do not fare as well as other undergraduates in the recently announced 1991–92 budget from the UFC (see opposite).

While the UFC has announced a 7 per cent increase in funding for all undergraduate teaching over 1990–91, the number of students in the physical sciences will grow by only 5.4 per cent, and fewer agricultural scientists, engineers and technologists will be funded than in the previous year. But biological sciences, with a 14.1 per cent increase, has bucked the trend.

The arts and humanities are expanded to a greater extent. Sir Peter Swinnerton-Dyer, UFC chief executive, says that the decisions were made "on the basis of apparent student demand".

The allocations also show that the UFC's research money is now being distributed more selectively, becoming concentrated in those universities which achieve the best research records. Swinnerton-Dyer told the House of Lords Science and Technology Committee last week that not all of Britain's universities can claim to be important centres for research, and the UFC can no longer afford to fund them all as such.

The idea of "teaching-only" universities was suggested by the Advisory Board for the Research Councils in 1987, but was eventually dropped from the government's plans to reform higher education in 1988 after protest from the universities. The UFC says that teaching-only universities are not on its immediate agenda, but goes on to add that individual universities may wish to spend their research money more selectively, and concentrate on teaching alone in specific subjects.

John Ashworth, director of the London School of Economics, says that separating teaching funds from research is "administratively convenient", but he argues that research spending in the universities helps to maintain teaching quality, because research students and postdoctoral researchers are involved in teaching.

The allocations do contain one piece of welcome news for university science. Last month, the UFC told the universities that it is planning to spend £70 million over three years to upgrade university animal facilities to make them comply with tough new standards for laboratory animal housing introduced in 1989, unexpectedly meeting the universities' requests for money almost in full (see *Nature* 345, 755; 28 June 1990). The universities will receive £35 million in 1991–92. John Bleby, from the Royal Veterinary College, London, says that the new money is "a terrific break for the animals", as well as for biomedical research.

Peter Aldhous

Lords look for culprits

London

THE UK House of Lords Science and Technology (S&T) Committee, which is investigating the financial crisis that followed the announcement of the 1991-92 science budget, will find no shortage of people to blame, if its first hearing is anything to go by.

The S&T committee aims to discover why the Science and Engineering Research Council (SERC) was forced to make swingeing cuts across its research programme after his year's budget was announced (see *Nature* 349, 551; 14 February 1991). But Sir Peter Swinnerton-Dyer, chief executive of the Universities Funding Council (UFC), last week told the S&T committee that SERC's current problems cannot be considered simply in the context of this year's budget, and accused SERC's previous chairman, Sir William Mitchell, of overstressing the council's resources by committing SERC to a series of expensive, long-term projects.

SERC's "day of reckoning" is the end result of a long-running crisis in British science, which is one of overdemand as well as underfunding, Swinnerton-Dyer suggested. In Britain, there is more research "waiting to be done" than any government could be expected to fund, he said, but towards the end of Mitchell's five-year reign, SERC failed to recognize this. British science administrators need to manage the country's research priorities more efficiently, according to Swinnerton-Dyer. Relying on increased funds from next year's science budget to cover future spending is "no longer an acceptable frame of thought".

SERC's current efforts to trim its programme, and to review the balance of its spending between "big science" facilities and small project grants, suggest that SERC is now actively reconsidering its priorities in the light of limited government funding. (The decisions announced so far, however, have met with widespread protest, notably the threatened closure of the Nuclear Structure Facility at SERC's Daresbury Laboratory.)

But Swinnerton-Dyer made clear his dissatisfaction with the UFC's absence from the main forum of debate over the science budget, the Advisory Board for the Research Councils (ABRC). As chairman of the UFC's predecessor, the University Grants Committee, Swinnerton-Dyer sat on the ABRC, but UFC chairman Lord Chilver insisted in 1989 that the new body should divorce itself from British research planning.

The ABRC's advice on research council funding affects the universities, but "the UFC has virtually not been consulted at all", over SERC's current difficulties, Swinnerton-Dyer told the S&T committee. Because few savings on large projects are possible this year, the vast majority of SERC's cuts in 1991-92 have fallen on project grants and research studentships, most of which are

awarded to university researchers.

SERC's chairman, Sir Mark Richmond, last week defended SERC against the accusations of lax financial management in the past, first levelled by Education and Science Secretary Kenneth Clarke. He told the S&T committee that the council's assumption that it would be given £40 million of new money for 1991-92 was "not wildly unreasonable" — an identical sum had been allocated for 1989-90. But he conceded that the £40 million figure may have been larger than warranted by the worsening state of the British economy, and the consequent tightening of the government's purse strings. In the event, SERC receive only £12 million above its "baseline" funding.

Swinnerton-Dyer also told the S&T committee that he is worried about the planned transfer of funds from the UFC to the research councils to cover the indirect costs of research council funded projects in the universities. This transfer begins with some £50 million in 1992-93. Swinnerton-Dyer said that SERC will be "under extraordinary pressure" to use the money to ease its own financial difficulties, rather than returning it to the universities to cover research overheads.

Peter Aldhous

MEDICAL RESEARCH

World's first MS unit

London

THE Multiple Sclerosis Society, a British charity, announced last week that it will spend nearly £2 million to set up the world's first research unit that will combine basic and clinical science to investigate multiple sclerosis (MS), the most common disease of the central nervous system.

Alan Davison, chairman of the society's research advisory committee, says the move was prompted by the decline in UK government support for medical research. These cuts, he says, mean that researchers in universities "now have to divert more of their time to teaching and clinical work".

In response, the medical research charities have taken over an increasing proportion of British medical research funding over the past decade, and now spend more each year in the universities than the Medical Research Council (MRC). And charitable societies are increasingly considering funding their own institutes, says Diana Garnham from the Association of Medical Research Charities (AMRC). This has been particularly evident in cancer research.

The new MS unit, which Davison hopes will open before the end of the year, aims to draw together the many strands of research into the disease, including immunology, virology and epidemiology, as well as neuroscience. The site has yet to be chosen.

Peter Aldhous

Moneysaving measures

London

LAST week, SERC chairman Sir Mark Richmond offered an appealing method for avoiding some of the cash crises that often plague the British funding of science. Speaking before the House of Lords Science and Technology (S&T) Committee, Richmond suggested that SERC be allowed to carry over as much as five per cent of its budget from one year to the next. In this way, the council could save money in years of plenty to use in leaner times.

SERC's problems, Richmond said, stem from the variability in government funding for science from year to year. In 1991-92, as much as £28 million of SERC's predicted shortfall is due to the failure of the science budget to keep pace with inflation and salary increases. As most of SERC's financial commitments run over a number of years, it would be "very helpful" if SERC could smooth out the variability in government funding, Richmond said. SERC is already allowed to carry over some of its budget from one year to the next, but only two per cent. Richmond wants that increased to five per cent of the annual budget, which would give SERC about £20 million to tinker with.

One potential problem with such a plan is that the Treasury may be tempted to reduce the science budget further if the largest research council is not spending all its funds in years of relative prosperity. Similar fears plague Japanese research spending, resulting in a profligate "end of year splurge" to spend research grants before the end of each fiscal year. Richmond's aides agree that is a risk, and say the issue must be addressed before SERC agrees to reforms, the Department of Education and Science has to argue the case with the Treasury, and SERC officials say no reforms are likely for a year or two.

Richmond also argued last week that subscriptions to international scientific programmes, which account for almost one third of SERC's spending, should be "ring-fenced", so that increases beyond SERC's control do not reduce spending on domestic science. Richmond would like to see some money provided from outside SERC's budget to pay for cost overruns caused by fluctuating currency exchange rates or inflation in other countries.

Like Sir Peter Swinnerton-Dyer (see left), Richmond believes the outlook of the Advisory Board for the Research Councils (ABRC), which advises the Education and Science Secretary on the science budget, is too narrow. "I'm not convinced", he said, that the ABRC takes enough account of the views of the Foreign Office which is responsible for many of the intergovernmental agreements governing SERC's contribution to international science projects.

Peter Aldhous

Wittgenstein's paranoia

SIR — The recent debate in *Nature* between Greenfield and Marshall¹ over the merits of Wittgenstein again shows that some people regard him as the greatest philosopher of the century, others as just one of a long line of German-speaking philosophers (such as Hegel and Heidegger) who have dazzled some innocent English-speaking philosophers by writing material whose basic nonsense is concealed by the impenetrable thickets of the German language. I would like to advance a third hypothesis. Russell in his autobiography (among others) draws attention to Wittgenstein's schizoid and paranoid personality. Canon C. E. Raven (who was vice-chancellor of the University of Cambridge during Wittgenstein's tenure of his chair there) told me that Wittgenstein at times actually suffered from paranoid delusions and would flee to outlying villages to escape from his imaginary enemies.

Certain schizoid personalities develop the ability to write in a form of speech disorder known as schizophrenese. In some cases this is by no means nonsense, but can have a powerful emotional and aesthetic appeal — it is in fact a kind of poetry, based on powerful allegorical and metalogical (not *logical*) associations. A well-known example in literature is *Finnegan's Wake*. Joyce himself was never overtly schizophrenic, although very schizoid, and his daughter developed the illness; but he must have been near enough to it to be able to write schizophrenese (which normal people find almost impossible to do). One essence of schizophrenese is that the meaning of a statement is never quite contained within the statement but lies somehow 'behind' it and when searched for continually retreats behind any further elucidating statements — a phenomenon that Wittgenstein's philosophical writings exhibit to a singular degree. In his book on Wittgenstein², Ayer, with his clear logical mind, makes attempt after attempt to understand what Wittgenstein is getting at — and continually fails, as he himself admits. I suggest that this is because Wittgenstein should be read as poetry and not as the kind of philosophy practised by Ayer and his intellectual forebears in the British empiricist tradition.

Wittgenstein's influence was also due to his remarkable hypnotic, even demonic, personality — Ayer told me that all the philosophers at Cambridge, except Broad, were terrified of him — and to his ferocious and overbearing style of philosophical debate. This is also documented by Teichman³. I had no personal contact with Wittgenstein (although my cousin Yorick Smythies was one of the Inner Circle), but when I was a research student at the Psychological Laboratory at Cambridge in the 1950s I used to attend the weekly meetings of Wittgenstein's disciples. I found myself unable to make any sense of the tortuous, not to say tormented, wrastlings with the thoughts produced by in-

spired members of the group in imitation of the Master's style. But, as a psychiatrist, I found them very like the thoughts and modes of thinking that troubled my schizophrenic patients.

Lastly, Wittgenstein's method was essentially based on an analysis of what he regarded as 'ordinary English usage'. Churchland⁴ has recently subjected this method to a devastating attack on the grounds that such 'folk psychology' is very unlikely to tell us anything true or interesting about the mind or about perception. One sympathizes with philosophers in their desperate defence of their turf against the continual inroads made upon it by the advances of neuroscience, perceptual science, cognitive science and perhaps an extended physics^{5,6} (from which sources our only true understanding of the mind can come); but to base this defence on such as Wittgenstein cannot advance their cause.

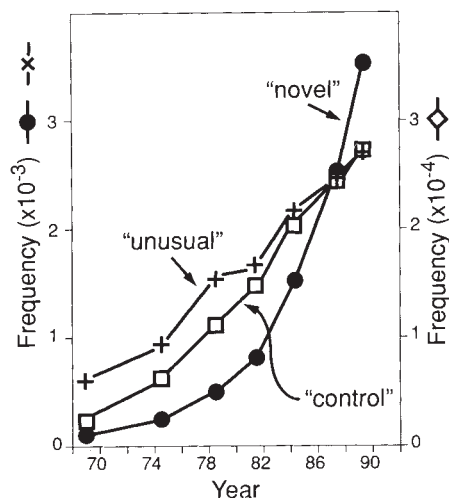
J. R. SMYTHIES

*The National Hospital
for Neurology and Neurosurgery,
Queen Square, London WC1N 3BG, UK*

1. Greenfield, A. J., Marshall, J. C. *Nature* **348**, 384 (1990).
2. Ayer, A. J. *Wittgenstein* (Weidenfeld & Nicholson, London, 1985).
3. Teichman, J. *Philosophy and the Mind*, 51 (Blackwell, Oxford, 1988).
4. Churchland, P. S. *Neurophilosophy* (MIT Press, Cambridge, 1986).
5. Penrose, R. *The Emperor's New Mind* (Oxford University Press, 1989).
6. Smythies, J. R. *The Walls of Plato's Cave* 6 (in the press).

So what's new?

SIR — A feature that distinguishes scientific journals from breakfast-cereal boxes is a preference for the word 'novel' over terms such as 'brand new'. This preference is accelerating, as is evident from the figure, where the frequency of 'novel' in titles and abstracts catalogued in Medline is compared to the frequency of a control word ('control')



over the past 25 years. The recent popularity of 'novel' is likely to be an inflationary increase rather than an explosion in the number of extraordinary discoveries. For one thing, the increase appears to come at the expense of less conspicuous synonyms, such as 'unusual', which shows a decline in popularity relative to 'control'. For another, 'novel' is especially favoured in journals that are noted more for solid science than for novelty. In the most recent month for which information is available, the *Journal of Biological Chemistry* and *FEBS Letters* each accounted for twice as many 'novel' papers (10) as the combined total for *Nature*, *Science* and *Cell*.

If this expanding use of 'novel' is allowed to continue unchecked, one can foresee that in the near future its presence on a paper will be *de rigueur*, and the word will lose its meaning entirely. To prevent this from happening, we recommend that authors reserve 'novel' for strikingly new discoveries and consider more appropriate terms to describe observations that range from somewhat atypical to unusual. We should also keep in mind that at the supermarket, novelties often end up at the rear of the ice-cream freezer.

STEVEN HENIKOFF*
ROBERT LEVIS

*Howard Hughes Medical Institute
and Basic Sciences Division,
Fred Hutchinson Cancer Research
Center,
Seattle,
Washington 98104, USA

New money

SIR — Whenever a new 'big science' mega-project is proposed, we are always reassured by the project's proponents that it will not adversely affect other areas of research because it will be financed with 'new money'. This term has been used so freely and so often that it is about time that someone asked precisely what 'new money' is supposed to mean. Surely the US government is not planning to finance the human genome project by printing 3,000 million new one dollar bills or the Superconducting Super Collider with 822,400 million newly minted pennies?

Perhaps George "read-my-lips" Bush will propose some more new taxes to produce this 'new money', or perhaps it will come from other non-science programmes such as food stamps or Medicare?

Physicists believe in the law of the conservation of energy. They would be the first to ridicule anyone claiming to have a perpetual motion machine that could produce 'new energy' out of nothing. Nevertheless, they expect us to believe *their* fable about 'new money'.

ROBERT J. YAES

University of Kentucky Medical Center,
Department of Radiation Medicine,
800 Rose Street,
Lexington, Kentucky 405036-0084, USA

Duesberg on AIDS and HIV

SIR—Pinching *et al.* cite¹ their unpublished cohort studies of HIV-positive and negative homosexuals and haemophiliacs as epidemiological proof for “a pathogenic role for HIV infection”, but, cohort studies are biased by those who recruit and by those who volunteer to be recruited. Before I could accept their conclusion, I would need to know for what “recreational drug use . . . the two groups were closely matched” and what AIDS-linked diseases they developed — whether, for example, Kaposi’s sarcoma was observed in homosexuals and haemophiliacs, or only in homosexuals. But it is surprising that, among 120 HIV-negative “sexually active homosexual men” and 15 HIV-negative “severe haemophilia A subjects”, not even one developed any of the 25 indicator diseases of AIDS in six years. Could it be that these diseases were not listed because in the absence of HIV they are called by their old names?

Further, the conclusion of Pinching *et al.* is difficult to reconcile with published epidemiological data. Since 1985, about a million people in the United States have been estimated by the Centers for Disease Control to be HIV-positive, yet not more than about 70,000 (7 per cent) have died and about 117,000 (12 per cent) have developed AIDS in the past 10 years². Moreover, about 15,000 people in the United States with severe haemophilia have been infected with HIV for about 10 years, the presumed average latent period for HIV. Yet during the past two years only 300, or fewer than 2 per cent, have died annually with a diagnosis of AIDS.

Thus US HIV-positives seem to have fared much better than Pinching *et al.*’s cohorts. Indeed, recent controlled studies have shown that HIV-free haemophiliacs have the same immunodeficiencies as HIV-positive controls and at the same rate^{3,4}, and that HIV-free male homosexuals have the same Kaposi’s sarcomas and some of the same T-cell deficiencies as HIV-positive controls⁵.

Apart from that . . .

SIR—I was dismayed to read the letter on AIDS by A. Karpas (*Nature* 348, 578; 1990). The speculation contained therein is unfounded and unnecessary. The letter is a series of *non sequiturs*. Such ramblings are nonsensical and possibly hurtful. The decision to submit and to publish this letter is reprehensible.

WILLIAM A. HASELTINE

Division of Human Retrovirology,
Dana-Farber Cancer Institute,
Harvard Medical School,
44 Binney Street,
Boston, Massachusetts 02115, USA

This, in my opinion, is why it is necessary to conduct randomized, controlled studies, instead of studying selected cohorts, to determine whether health risk factors, such as “recreational drugs” and “severe haemophilia” can cause AIDS without HIV, and whether HIV can cause AIDS without such factors.

PETER H. DUESBERG

University of California, Berkeley,
Department of Molecular and Cell
Biology,
c/o Stanley/Donner Administrative
Services Unit,
Berkeley,
California 94720, USA

1. Weiss, R. A. & Jaffe, H. W. *Nature* 345, 659–660 (1990).
2. Duesberg, P. H. *Nature* 346, 788 (1990).
3. Pinching, A. J., Jeffries, D. J., Harris, J. R. W., Swirsky, D. & Weber, J. N. *Nature* 347, 24 (1990).
4. US Dept of Health and Human Services, Centers for Disease Control HIV/AIDS Surveillance (January 1990).
5. Jin, Z., Cleveland, R. P. & Kaufman, D. B. *J. Allergy clin. Immun.* 83, 165–170 (1989).
6. Madhok, R. *et al. Br. med. J.* 293, 978–980 (1986).
7. Friedman-Kien, A. E. *et al. Lancet* 335, 168–169 (1990).
8. Marlon, S. A. J. *Acquired Immune Deficiency Syndrome* 2, 178–186 (1989).

Feel the draft

SIR—Your analysis of “Conceptual sovereignty problems” (*Nature* 348, 467; 1990) is flawed at the outset by an historical error: “The monarch . . . would conscript an army of serfs . . .”.

Monarchy, at least in the Middle Ages and in modern times in Europe, has never known conscription. In the Middle Ages, armies consisted of noblemen, vassals called to service by their suzerain; a king could call the vassals of his vassals (“*l’arrière-ban*”). Serfs were specifically excluded from military service, as had been slaves in antiquity. When an army of knights came to show its inefficiency, at Agincourt and other places, it was progressively replaced by professional armies with long-term, volunteer service. The last time the King of France called the “*arrière-ban*” was in the black year of 1709. To this day, Britons, having the privilege of being ruled by a sovereign, are immune from conscription except in desperate circumstances.

Conscription was invented by the French Revolution from necessity, France having declared war on Europe and most noblemen having emigrated or had their heads cut off. Another factor was reference to the “democracies” or republics of antiquity, where an army of conscripted free men (that is, slave owners) was necessary, among other things, to keep slaves from overthrowing their masters.

In fact, your argument could be correct in a roundabout way. ‘Democracies’ (whatever this much-abused word may mean) seem constitutionally incapable of

peacefully coordinated action, not to speak of federation (it took the Sonderbund war to create model, modern Switzerland, and the ugly war between the states really to unite the United States). The nearest thing to unity Europe has known since the fall of the Roman Empire had been created by the conscripted armies of that heir to the French Revolution, Napoleon; mercifully, it did not last, any more than did the empire created by Stalin’s (also conscripted) armies. Only when identity and sovereignty were both subsumed in a monarch (or, less well, in a proud elite of burghers) have their conflicts been resolved. The modern solution, alas, seems to be force and conquest: shotgun federation in the true sense of the word.

J.-F. FONCIN

Laboratoire Montyon-La Salpêtrière,
47 Boulevard de l’Hopital,
75651 Paris Cedex 13,
France

Radiation doses

SIR—There is no sense in the reduction of permissible radiation dose limits recommended by the International Commission on Radiological Protection as reported by Peter Aldhous and Seth Shulman (*Nature* 348, 274; 1990).

The long-running studies reporting a three times higher risk to irradiated survivors of Hiroshima and Nagasaki have been concerned with the more heavily irradiated individuals and a linear extrapolation of risks to doses in the millisievert range; this is no more valid than to suppose that consumption of a single aspirin tablet must lead to a half per cent risk of death because 200 aspirins taken together would be reliably lethal.

Among the Japanese bomb survivors, no observable excesses of cancer deaths were recorded for those receiving less than 50 rem (0.5 Sv).

In my view, the conclusion of the Gardner report, which linked leukaemia in children with the radiation exposure of their fathers, has many arguments against it. For example, the natural background in Kerala in India and in parts of Brazil equalled the dose recorded for the fathers quoted by Gardner, with no corresponding increase in childhood leukaemias among the children of visitors to those areas.

I find it disturbing that so much more attention is constantly given to the very improbable risks of low radiation doses than to the real dangers of air pollution by vehicle exhausts and coal burning. Such air pollution is probably responsible for more than 10,000 deaths a year in Britain alone.

J. H. FREMLIN

46 Vernon Road, Edgbaston,
Birmingham B16 9SH, UK

Environmental impact of fires in Kuwait

Richard D. Small

The deliberate firing of the Kuwaiti oil wells by the Iraqis is an act of gross environmental vandalism. But the likely impacts on the climate have been exaggerated.

THE Iraqi invasion of Kuwait on 2 August 1990 set in motion actions threatening the environment on a grand scale. Iraq demonstrated its intent to use environmental destruction as an instrument of national policy. Several million barrels of oil have been released into the Persian Gulf; several hundred oil-well fires have been started. Smoke clouds have covered Kuwait and spread over Iran reaching toward Pakistan. Black rain has fallen over Kuwait, the Gulf and Iran from Qasr-E-Shirin in the north to Bushehr in the south, adversely affecting population, water supply and agriculture.

Anticipating these actions, scientists met in London in January to consider possible climatic consequences of the Gulf War¹. Some assessed the potential for environmental damage as minimal; others forecast catastrophe. The more extreme predictions described substantial surface cooling, failure of the Asian Monsoons, extensive crop failures and accelerated global warming.

Here, drawing on diverse disciplines including reservoir dynamics, combustion, fire plume dynamics, aerosol dispersion, smoke physics and climate impacts, I apply models to estimate smoke production and assess environmental impacts. Although assumptions have to be made, a worst case can be identified and assessments made with confidence. My conclusion is that, although the firing of the oil wells represents a major pollution event with unknown ecological consequences, the likely impacts on the climate have been exaggerated.

In 1989 Kuwait produced 1,593,000 barrels of oil per day ($B d^{-1}$) in seven fields². Kuwait's three refineries are integrated with crude oil and products piped between them. An estimated 13 million barrels were stored before 2 August 1990. The principal fields and refineries are shown in the figure.

Catastrophe

Destruction of the wells will cause fires that are likely to continue until extinguished. They are isolated 'point' fires supported by the natural flow rate of each well. The fires would not merge (the wells are 800–1,600 m apart), but until the pressure is reduced, they will continuously add smoke to the atmosphere. The refinery fires will be difficult to extinguish, but the burning and smoke production is limited to the stock at hand. Because of the natural flow rates in Kuwaiti wells and the refinery inventory presumably still on hand, the potential exists for an environmental catastrophe.

I develop the calculation so as to estimate

the maximum amount of smoke likely to be produced from the burning oil wells and hydrocarbon storage at the refineries. The 1989 average daily production rate (rates are highly variable) was $1,593,000 B d^{-1}$ with an average flow rate per well of $4,390 B d^{-1}$. That flow was maintained by injection of natural gas (purchased from Iraq!); and is greater than the average 1989 Kuwaiti well production of $3,427 B d^{-1}$ estimated by Al-Marafie³. The rate per well is determined by the reservoir pressure and the number of producing wells. As more wells are brought on stream, the rate falls^{3,4}.

The fields have not been operating since early August and gas injection has been halted. Reservoir pressures are thus frozen at the early August levels. Immediately following destruction of the well heads, oil would flow at the last production rate. The reservoir pressure will fall, and the oil flow diminish with time. The decline could occur over months or years depending on the reservoir structure and the well location⁴. I assume a constant flow and neglect any decrease as the reservoir declines.

The amount of smoke produced is calculated from the product of the well flow rate P ($B d^{-1}$), crude density m ($kg B^{-1}$), and smoke

emission factor. Values for each parameter are summarized in the Table. I assume that the entire production rate is completely burned each day. Before the invasion, 363 wells were in use. By blowing the shut-in wells, the average flow per well would be reduced; the total crude flow would not be greatly increased and, without injection, possibly reduced. Nevertheless, for the purposes of a worst-case analysis I assume that a 25 per cent increase in flow is possible.

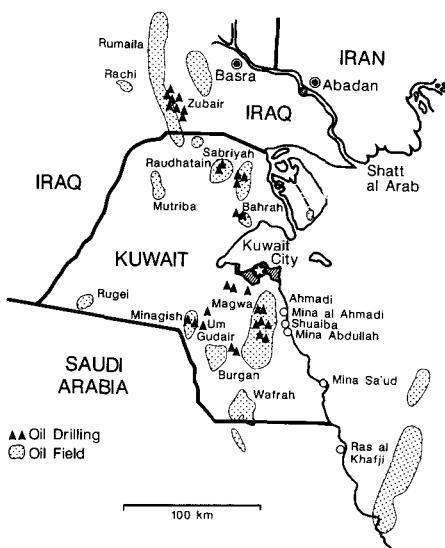
Most of the crude is drawn from the Burgan field ($^{\circ}API$ gravity 30.8; specific gravity 0.87). It is slightly different for the other fields; the variation is small. Based on the average flow rate per well I calculate a crude exit velocity of $2 gal s^{-1}$ or $1 m s^{-1}$. Such flows are not extreme, but they are nonetheless substantial. Flow rates in individual wells can increase or decrease by factors of 2–3.

Smoke emission factors for crude oil (determined from laboratory experiments) range from $20 g kg^{-1}$ to $98 g kg^{-1}$ with an average value of $73 g kg^{-1} \pm 25 g kg^{-1}$ (ref. 5). Roughly 69 per cent of the emission is elemental carbon. Such smoke is high absorbing (albedo = 0.3, extinction coefficient = $10.6 m^2 g^{-1} \pm 1.4$) in the visible wavelengths. Using the mean emission factor, the amount of smoke produced per well is $S = 4.4 \times 10^7 g d^{-1}$, and for the entire production of $1,593,000 B d^{-1}$, the smoke emission is: $S_{wells} = 0.016 Tg d^{-1}$ where one terragram (Tg) is one million metric tons. In a 30-day period, 0.48 Tg of smoke could be added to the atmosphere. That mass would be supplemented by smoke from the refinery complexes should they also be fired.

Because the inventory at the refineries contains both crude and products, the smoke emission factor is lower. For a typical product mix⁵, the weighted emission factor is $50 g kg^{-1}$. Assuming an inventory of 32 per cent crude and 68 per cent products ($^{\circ}API$ gravity = 45), the weighted emission factor is $57 g kg^{-1}$ and the potential smoke production ($S_{refineries}$) is 0.10 Tg.

This is six times greater than the amount of smoke produced by all the burning wells in one day. Moreover, it is possible that the entire smoke emission would take place within 5–10 days. Thus I estimate a possible smoke emission of $S = 0.016 + 0.010 = 0.026 Tg d^{-1}$ for 10 days and $S = 0.016 Tg d^{-1}$ for each day thereafter until the well fires are extinguished. In the first month the smoke emission could total 0.58 Tg.

The potential for environmental damage depends on both the smoke mass and the injection height. In general, smoke from the



Kuwaiti oil facilities. Over 80 per cent of the production is from the Burgan (210) wells and Magwa (71 wells) fields. The refineries at Mina Abdullah, Mina al Ahmadi and Shuaiba have a combined capacity of $670,000 B d^{-1}$. The 13,000,000 B storage at the refineries corresponds roughly to a 10-day inventory of crude and products.

SMOKE CALCULATION PARAMETERS FOR FIRES IN KUWAITI OIL FACILITIES

Category	P 10^6 B d^{-1}	$^\circ\text{API}$	m kg B^{-1}	τ g kg^{-1}	S Tg d^{-1}
Well-Head Fires*	1.6	31.6 $\pm$ 4.7	138 $\pm$ 4	73 $\pm$ 25	0.016 $\pm$ 0.006
Maximum flow [†]	2.0	31.6	138	73	0.020
Refineries [‡]					
Crude	0.43 [§]	30.8	138	73 $\pm$ 25	0.004 $\pm$ 0.0005
Products	0.88 [§]	45.0	127	50 $\pm$ 15	0.006 $\pm$ 0.001
Total	1.31.31 [§]	40.5	130	57 $\pm$ 19	0.010 $\pm$ 0.003

*August 1990 production (ref. 2).

[†] See refs 3, 4 for production tradeoffs.

[‡] Normal operating inventory assumed; current inventory is probably lower.

[§] Per day consumption assuming inventory burns for ten days.

^{||} $^\circ\text{API}$ gravity for Burgan crude.

refinery fires will be injected higher than smoke from the burning wells. This follows because a burning storage tank has a larger plume cross-section or smaller aspect ratio than a well-head fire. As a consequence, the buoyancy is not as rapidly diffused by mixing of 'cold' ambient air and the plume rises higher in the atmosphere.

Hydrocode solutions for very large oil-pool fires show maximum smoke injection altitudes of roughly 3 km (refs 6,7). Most of the smoke spreads out at 2 km—well below the top of the cloud. This is for a quiescent atmosphere; ambient winds reduce the rise. Heikes *et al.*⁸ showed similar smoke injections for comparable size area fires in moist atmospheres. In desert conditions, the smoke rise should be between 1–3 km.

Smoke from the burning wells will not be injected as high as the smoke from the refineries because the plume is slender and the buoyancy more rapidly diffused. The initial velocity at the well head only slightly adds to that generated by the heat release ($\sim 45 \text{ KJ g}^{-1} = 319 \text{ MW}$). For a flame temperature of $\sim 1,000^\circ\text{C}$ and an initial well-head velocity of 1 m s^{-1} , the maximum plume centreline velocity is $\sim 7.2 \text{ m s}^{-1}$. This corresponds to a smoke injection height of roughly 700 m. Higher injections (to 1 km) are possible as are lower injections.

In summary, I expect a bimodal smoke injection. Smoke from burning refinery stocks could reach 1–3 km, whereas smoke from the oil fields generally will stay below 1 km. Observation of smoke rise from the refinery and oil-well fires in Kuwait show that calculated worst case injections are overestimating the plume rise.

Impacts

Smoke is a strong absorber of solar radiation. A large smoke mass in the atmosphere can lead to a reduction in solar radiation reaching the Earth's surface and cause a cooling in the lower atmosphere. Global climate calculations show that only minimal climate changes occur for small absorption optical depths ($\tau_a \ll 1$).

The atmosphere can be affected on several scales. At the local scale (10^{11} m^2), the smoke concentrations are greatest and radiative heating of the surface most effected. As the smoke spreads over regional (10^{13} m^2) and global scales (10^{14} m^2), the clouds thin, but weather or climate are possibly

influenced. A number of impacts have been asserted. They include smoke clouds covering 20 per cent of the Northern Hemisphere, cooling of the Earth's surface similar to a 'nuclear winter', failure of the Asian monsoons, and heating of the atmosphere due to increased CO_2 production. Such consequences are serious and should be neither advocated nor dismissed without cause.

Assuming the continuing smoke production is dispersed at a mean wind speed of $\sim 5 \text{ m s}^{-1}$ (consistent with smoke stabilizing at 2 km), an area coverage of $15 \times 10^{12} \text{ m}^2$, an absorption coefficient for the smoke mixture of $6.5 \text{ m}^2 \text{ g}^{-1}$ (ref. 9) and a smoke fallout rate of 80 per cent (appropriate for an injection to 2 km), the absorption optical depth at the end of one month would be:

$$\tau_a = \frac{k_a S}{A} = \frac{6.5 \text{ m}^2/\text{g} \times (1.0 - 0.80) \times 0.58 \text{ Tg}}{15 \times 10^{12} \text{ m}^2} = 0.05$$

This optical depth reduces solar radiation to the surface only slightly and little temperature change is expected¹⁰. In fact, because the smoke remains at low altitudes, a slight warming may result¹¹. Even though radiative heating of the ground is reduced, the lower atmosphere warms as the smoke layer absorbs the solar radiation. Balancing the reduced heating is a decreased cooling of the surface by convection and an increased downward infrared radiation from the hotter atmosphere. The net effect is a decrease in peak daytime temperatures, but an increase in the daily-average surface temperature.

The estimate of τ_a is probably high as I assume a particle longevity not justified by the low injection heights (assuming an eddy viscosity of $7 \text{ m}^2 \text{ s}^{-1}$ and an injection height of 2 km, the e-folding time or mean lifetime of smoke is 7 days), and I assume that all hydrocarbons are completely burned. Moreover, I assume that in sabotaging the wells, damage to the piping does not restrict the flow. Furthermore, I allow that the crude flow would be maintained despite decline of reservoir pressures. Some decline is already apparent. Smoke lifetimes are considerably reduced for lower-altitude injections. For well-head fires (1-km maximum altitude), the e-folding time is roughly 40 h; and the dispersion reduced considerably. The environmental cost would be an increased local fallout. The cloud patterns observed to date show that the smoke lifetime is less than that used in the worst-case estimate.

I have also estimated the CO_2 addition to the atmosphere. Refinery fires would contribute at most 1.3 Tg; well-head fires would produce 0.17 Tg d^{-1} . In one year a total of 63 Tg of CO_2 could be produced. This is roughly one per cent of the global CO_2 production from fossil fuels and increases the total CO_2 content of the atmosphere by 0.0025 per cent. No warming would result.

Even though the wells may continue to burn and produce smoke ($\sim 0.48 \text{ Tg}$ per month) for a prolonged period, the dispersion is likely to be regional rather than continental because the injection is so low. Most of the fallout will probably occur over the Persian Gulf, southern Iran, Pakistan and in the extreme northern India.

Conclusions

My estimates of the smoke mass produced by destruction of Kuwait's oil wells and refineries and the smoke stabilization altitude do not support any of the purported climate impacts. The smoke is not injected high enough to spread over large areas of the Northern Hemisphere, nor is enough produced to cause a measurable temperature change or failure of the monsoons. Similarly, only a small increase in the global CO_2 budget results.

Nevertheless, the amount of smoke produced would be large and cause a massive pollution event. It would impact the ecology of the Persian Gulf and fall out on a wide swath across southern Iran, Pakistan, and possibly northern India. Because the fires could continue for a prolonged period, the particulate fallout would persist. In the worst case, human population would be exposed to persistent air pollution; grazing herds may be affected by accumulated ingestion of soot; agricultural and ecological systems impacted by a covering of soot; and water systems polluted by the fallout and accumulations in runoff. The impact on human and animal populations and regional ecosystems from such prolonged effects is unknown. □

Richard D. Small is at the Pacific-Sierra Research Corporation, 12340 Santa Monica Boulevard, Los Angeles, California 90025, USA.

1. Aldhous, P. *Nature* **349**, 96 (1991).
2. *Oil Gas J.* **88**, 46–83 (1990).
3. Al-Marafie, A. M. R. *Energy* **14**, 1–7 (1989).
4. Sheshodia, T. S., Mubarak, S. I., Javed, M. S. & Rahman, B. A. in *Proceedings: 6th Middle East Oil Show*. Soc. of Petroleum Engrs. of AIME, 285–301 (1989).
5. Ransohoff, L. M., Knudson, K., Bush, B. W. & Small, R. D. *Material Inventories and Smoke Properties for U.S. Target Areas* (Pacific-Sierra Research Corporation, 1989).
6. Panner, J., Haselman, L. Jr & Edwards, L. L. *J. Clim. Appl. Meteor.* **25**, 1434–1444 (1986).
7. Briggs, G. A. in *Lectures on Air Pollution and Environmental Impact Analyses* (ed. Haugen, D.A.) (American Meteorological Society, Boston, 1975).
8. Heikes, K. E., Ransohoff, L. M. & Small, R. D. *Atmos. Environ.* **24A**, 297–307 (1990).
9. Small, R. D. *Ambio* **19**, 377–383 (1989).
10. Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. *Science* **247**, 166–175 (1990).
11. Cess, R. D., Potter, G. L., Ghan, S. J. & Gates, W. L. *J. Geophys. Res.* **90**, 12937–12950 (1985).
12. Post, W. M., Peng, T., Emanuel, W. R., King, A. W., Dale, V. H. & DeAngelis, D. L. *Am. Sc.* **78**, 310–326 (1990).

A dose of vitamins and a pinch of salt

The study, published last week, purporting to show that children's IQ can be increased by dietary supplements, is an unhappy piece of work. The association could well be due to chance.

Does intellectual power grow from the neck of a bottle of pills? With all the ballyhoo about the effect of dietary supplements on IQ — discussed also on pages 2 and 5 of this issue, and in the main mercifully so far confined to Britain — one might be forgiven for supposing that it does. The scientific controversy centres on a "symposium" issue of the journal *Personality and Individual Differences* (12, 329–366; 1991), which contains studies by S. J. Schoenthaler, H. J. Eysenck, J. Yudkin and colleagues purporting to show that a modest dietary supplement of 23 vitamins and minerals taken on a sustained basis leads to a significant elevation of non-verbal IQ.

The specific importance of non-verbal tests in this context is that the abilities they tap are thought to be less affected by schooling and to be a better reflection of a person's underlying potential than those represented in verbal tests. The claimed improvement is 3.7 points using the well-known WISC-R test over a 13-week experimental period. The effect is reported only for a supplement of 100 per cent of recommended daily allowance (RDA), however, not for groups given 50 per cent or 200 per cent of RDA, and no claim is made for improvements in verbal IQ or full-scale IQ. The RDAs in question, of 12 vitamins and six minerals, are set at the US levels, rather than the lower UK levels. For four further minerals included in the treatment, and for vitamin K, no RDAs have been established.

Of principal interest is the large-scale study by Schoenthaler *et al.* which involved three treatment groups and a placebo group in a double-blind design. Twenty-nine scores on cognitive tests of one kind or another were analysed, but three of these (possibly five — the text is not clear) are summary scores, and not independent of the others. Three contrasts (each treatment group with placebo) were available for each test, resulting in a total of 87 contrasts. Seven of these yielded "significant" results — the object assembly and coding subtests of WISC-R and the non-verbal IQ score to which they contribute, the arithmetic subtest (which contributes to verbal IQ), and the reading and comprehension tests from the Comprehensive Test of Basic Skills, plus the total score from this battery.

Other non-verbal tests, including three further subscales of WISC-R, Raven's Matrices and the Matrix Analogies Test, produced no significant effect, nor did five other verbal subtests of WISC-R. That is to say, four of the seven effects reported to be significant

are related to verbal IQ or achievement tests, three to non-verbal IQ; looking only at scores that are independent of each other, five out of 78 contrasts are significant, of which three relate to verbal IQ, two to non-verbal. By chance, about one result in 20 is expected to reach the criterion for significance: given the tendency of scores on these tests to intercorrelate highly, the results are too close to what might be expected by chance to inspire confidence in the findings.

The results, however, are not presented in quite this way. Some of the time significance tests are not presented at all, but are replaced by rank orderings of average scores in the various groups, with a passing reference to significance tests in the text. There is no overall multivariate analysis of variance done, which, because the various test scores may be expected to be heavily intercorrelated, severely limits the interpretability of results. Instead, many univariate statistical tests are applied, leaving the sense that the data have been quarried for support for the desired effect rather than analysed on a statistical model. There is the reek of conviction driving analysis and presentation, of Schoenthaler *et al.* striving to make a case, rather than coolly analysing experimental results. Where results do not fit the hypothesis, special pleading comes to the rescue. For instance:

The absence of a significant relationship using the MAT [Matrix Analogies Test — a non-verbal test] was not totally unexpected. Since the test is given in a group setting, the opportunity for students to take the test lightly and rush through it was present. The large standard deviation suggests a number of students did just that. However, the trend between the 4 groups was the same as found on the nonverbal half of the WISC-R. More specifically, the placebo group did the worst and the 100% RDA group the best, making the MAT supportive of the main findings.

A more blatant piece of post-hocery is scarcely to be found. It is one thing to analyse data every which way to search for hypotheses for testing on a replication sample, quite another to present several forms of analysis chosen to suit a preferred hypothesis as proof of that hypothesis.

There are plenty of other holes in the information presented. For instance the precise basis on which raw scores were converted to IQs — important because WISC-R uses age-based norms, and in their teenage years children are getting better at the skills tested — is not given; nor is information on actual test score statistics for the four experimental groups; nor are details of correlations amongst the various measures.

The results of assays of blood samples, taken as part of the study, were not available at the time this research was published (other work in the symposium issue suggests that any effect of dietary supplementation is reflected in blood nutrient concentrations). And the results of a parallel study undertaken in England (rather than California) were not ready. So why publish now? No front-rank journal would contemplate publication in this form of incomplete findings bearing on such an important issue.

This is not to say that there is no phenomenon worthy of attention. Other, smaller-scale work described in the same symposium suggests that vitamin and mineral deficiency may be an important determinant of IQ and behaviour amongst delinquent adolescents, and that remedying the deficiency has desirable outcomes. It is not difficult to believe that *some* children would benefit in their intellectual development from dietary supplements or better eating habits, and there are enough hints in the data consistent with this view to make it worth seeking means of identifying who they may be. But the recommendation made by Eysenck, that the best policy would be to administer a 100 per cent RDA dose to all children on the basis of the present evidence is at best premature. What, for instance, of those children who did worse than expected in the treatment groups?

Schoenthaler *et al.* interpret their results as showing that dietary supplementation improved non-verbal intelligence estimates by a minimum of 6 points, an average of 11 points and a maximum of 21 points. This is so only for subjects preselected for showing large gains in scores over the experimental period. This style of interpretation of difference scores is psychometrically naive and thoroughly misleading. The authors write: "no change is expected in most students with dramatic changes appearing in others", but no evidence is presented of such a marked bimodality in the data.

The purported significant elevation of IQ as a result of vitamin treatment is a feeble thing at best — it is neither robust across different tests, nor large, and the claim that it relates to non-verbal rather than verbal tests is not supported by the data. Parents who buy dietary supplements because they have been told they will make their children more intelligent are, on this evidence, being misled.

Steve Blinkhorn

Steve Blinkhorn is managing director of Psychometric Research and Development Ltd, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK.

Flux lines on video holograms

David Caplin

THE elementary quantum of magnetic flux that appears in superconductivity is large enough that the penetration of magnetic fields into superconductors is sufficiently coarse-grained to be visible with electron microscopy. Pictures of the distribution of discrete flux lines were first obtained more than 20 years ago, but a new electron holographic technique developed by Tonomura and co-workers¹ at Hitachi's Advanced Research Laboratory in Japan allows the flux lines to be imaged dynamically, with an image rate of tens of frames per second. So far, the technique has been applied only to a conventional low-temperature superconductor, lead, but it should be usable also with technologically significant materials, such as the niobium alloys from which high-field magnets are fabricated, and perhaps eventually with the new high-temperature oxide superconductors. Because the crucial parameter of critical current density is controlled by the physical pinning of flux lines, such studies should help to improve their performance.

The flux quantum ϕ_0 , which is the ratio of Planck's constant, h , to twice the electron charge e ($2e$ because superconducting electrons are paired), is equal to 2.07×10^{-15} webers. In the mixed state of a type II superconductor immersed in a magnetic field, the field penetrates in the form of quantized flux lines at an areal density proportional to the magnetic induction B . If the latter is 1 millitesla, there are 5×10^{11} flux lines per square metre, corresponding to a typical spacing between lines of about $1 \mu\text{m}$, and so, well

within the resolution of electron microscopy.

The earlier method² of imaging flux lines was to form a Bitter pattern by evaporating a 'smoke' of submicrometre ferromagnetic particles onto the surface of the superconductor. The particles are attracted to regions of strong field gradient, which in this case are the points at which the flux lines emerge from the surface of the superconductor. Under appropriate conditions, in both conventional² and high-temperature^{3,4} superconductors, the flux lines have been seen to form a regular lattice.

The drawback of the Bitter method is that, because it requires the formation of a replica of the flux-line pattern, it cannot be used to follow the motion of the flux lines in the superconductor. The Hitachi group's electron-holographic technique overcomes this handicap, and so can distinguish between flux lines that are pinned firmly in place, and those that are comparatively mobile.

As in making any hologram, what is recorded in electron holography is the interference pattern between a reference wave and the object wave; the pattern depends on both the amplitude and phase of the latter wave. The Hitachi group, led by Akira Tonomura, has been developing electron holography for more than 20 years, and have applied it to a wide range of problems⁵. In the new work, it is the magnetic field of the fluxon that alters the phase of the electron wave. This is just the Aharonov-Bohm⁶ effect (once a theoretical curiosity in quantum mechanics), in which the induced phase difference, measured in wavelengths, between two paths (Fig. 1) is simply equal to the magnetic flux they enclose, measured in units of h/e . Thus a single flux quantum induces a phase difference of half a wavelength, and so causes destructive interference.

The big advance that has now been made by Matsuda *et al.* is to replace photographic recording of the hologram, which takes several seconds, with much faster video recording. Furthermore, instead of reconstructing

the photographic image optically with laser light, the more conventional approach, a fast computer is used to obtain the image by numerical Fourier transform algorithms. This allows the full range of image analysis and enhancement techniques to be deployed, which, because the shape of the fluxon field lines has to obey the laws of classical electromagnetism, are rather powerful.

In a structurally perfect superconductor, the current-induced Lorentz force on the

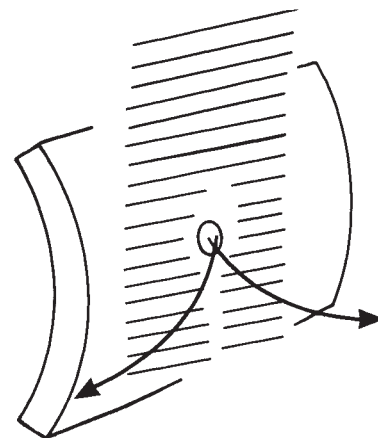


FIG. 1 A flux line trapped in a superconducting film induces phase changes in an electron wave. A single flux quantum, magnitude $h/2e$, gives a half-wavelength phase difference between the wave fronts on either side of it.

flux lines would cause them to move easily, dissipating energy, so that the critical current would be essentially zero. In conventional superconductors, dislocations and micrometre-scale precipitates are known to be effective at pinning flux, and so raising the critical current to useful levels. In the new oxide materials, there are good reasons to believe that only much smaller crystallographic defects can be effective; perhaps oxygen vacancies provide strong pinning sites⁷, but so far there is really no microscopic evidence to link the position of a pinned flux line to identifiable microscopic changes in structure.

The electron holographic technique may open the way toward microscopic correlation between structure and pinning capability. Simply because the length scales are larger, the task will be easier with the low-temperature niobium-based superconductors. The challenge of applying the technique to the high-temperature materials is daunting, but if successful, could provide the key to using them as practical conductors. □

David Caplin is in the Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BZ, UK.

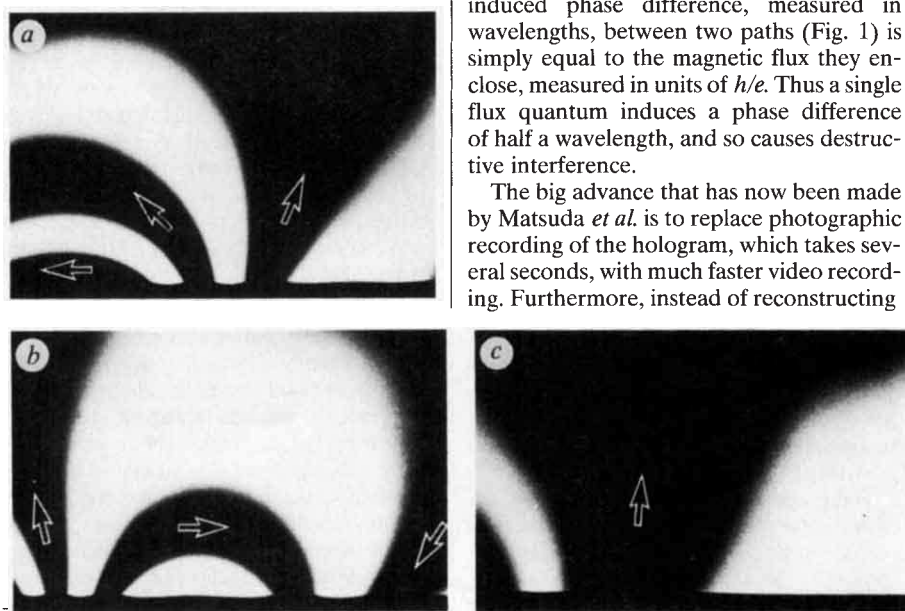


FIG. 2 Flux lines trapped in a superconducting lead film¹ at 5 K, only just below its superconducting transition temperature of 7 K, so that pinning is weak and the flux lines are highly mobile. Initially *a*, three 'up' flux lines are visible in the centre of the frame; *b*, 0.13 s later, two 'down' flux lines have entered from the right; and *c*, at 1.33 s, a pair of 'up' and 'down' flux lines have annihilated. (Courtesy A. Tonomura.)

1. Matsuda, T. *et al.* *Phys. Rev. Lett.* **66**, 457–460 (1991).
2. Essman, U. & Trauble, H. *Phys. Lett.* **24A**, 526–527 (1967).
3. Gammel, P.L. *et al.* *Phys. Rev. Lett.* **59**, 2592–2595 (1987).
4. Vinnikov, L. Ya. *et al.* *Solid State Comm.* **70**, 1145–1148 (1989).
5. Tonomura, A. *Phys. Today* **43**, 22–29 (April 1990).
6. Aharonov, Y. & Bohm, D. *Phys. Rev.* **115**, 485–491 (1959).
7. Daeumling, M., Seuntgens, J.M. & Larbalestier, D.C. *Nature* **346**, 332–335 (1990).

We may not have a morphogen

Jeremy Brockes

ATTENTION has centred on retinoic acid (RA) as a potential vertebrate morphogen because of its ability to alter axial specification in avian limb development and urodele limb regeneration (reviewed in refs 1–3). In the chick limb bud the pattern of mesenchymal tissues on the anteroposterior (AP) axis is specified by a mesenchymal cell group on the posterior margin, referred to as the polarizing region or zone of polarizing activity (ZPA). If this cell group is grafted to the anterior margin of a host bud, it establishes a second posterior boundary that results in formation of a mirror-image duplication of digits along the AP axis. The ability of a locally implanted source of RA to mimic such a graft led to suggestions that RA was produced by the ZPA, that its graded distribution across the bud specified the AP axis, and that it was thus the first identified vertebrate morphogen. Reports on pages 81 and 83 of this issue^{4,5} provide evidence for a significantly different interpretation — namely that a local source of RA induces neighbouring tissue to become a polarizing region, which then determines the AP axis by some mechanism not involving the graded activity of RA.

It has been proposed for some time that retinoic acid could act by inducing limb tissue to become a ZPA⁶. Wanek *et al.*⁴ report that after implantation of an RA-loaded bead in the anterior margin, it is possible to remove a wedge of tissue adjacent to the bead and demonstrate polarizing activity on grafting it to a recipient limb bud. Such wedges show activity after 16 hours of exposure to RA, but not after 12 hours, and induction is essentially complete after 24 hours. The chick-quail marker system has been used to demonstrate that the grafted cells do not make a significant contribution to the pattern duplication, and so are acting as a signalling centre.

Carryover hypothesis

It is important to rule out the possibility that the effect is due to carryover of RA in the grafted tissue. Wanek *et al.* have allowed a clearance interval between removing the bead and excising the wedge. Together with the failure of 12 hours exposure to induce activity and the known kinetics of RA removal and release from the bead, the authors argue that the carryover hypothesis is not valid. It might nonetheless be desirable to obtain direct evidence on this point from studies with labelled RA. Wanek *et al.* suggest that the graded response to exogenous RA may reflect variation in the number of adjacent cells induced to become polarizing cells. They point out that the polarizing activity could be unrelated to RA, or alternatively that RA acts in an autocatalytic relay whereby cells along the axis are

signalled to produce more RA, thereby retaining its status as a possible morphogen.

Noji *et al.*⁵ provide evidence against this second hypothesis, thus complementing the study of Wanek *et al.* These authors and others^{7,8} have identified a β -type retinoic acid receptor (RAR) expressed in the chick limb bud. The promoter of RAR β in human and mouse contains a retinoid response element that mediates a rapid elevation of RAR β transcription on application of RA^{9,10}, and the authors have used this effect as an endogenous reporter of RA activity in the distal bud. When a bead containing RA, its morphogenetically active relative 3,4-didehydro retinoic acid (ddRA)¹¹ or a synthetic retinoid is implanted on the anterior margin of the bud, the local induction of RAR β can be detected by *in situ* hybridization within four hours of implantation. Similar results were obtained after implantation of beads carrying synthetic retinoid into the polarizing region on the posterior margin. Despite the clear response to exogenous retinoid, a graft of ZPA tissue does not provoke any detectable increase in RAR β expression, and the authors conclude that RA is unlikely to be released by the ZPA either on the posterior margin or after grafting.

These results are provocative, but it is uncertain that the RAR β response is a completely satisfactory reporter. Rather little is known about intracellular retinoid traffic, or about the disposition of binding proteins, degradation systems and other components that might modify the responsiveness of the endogenous β gene. All this might conspire to make a functional distinction between what the authors refer to as “endogenous” retinoid released by the ZPA, and exogenous retinoid applied on a bead. A related possibility is that the ZPA delivers retinoid with a concentration profile that is not detected by the β response.

These possibilities do not seem particularly likely, but the study by Noji *et al.* will have to be confirmed and extended with other reporter systems before firm conclusions can be drawn from it. The authors have nonetheless established that a bead implant does not precisely mimic a ZPA graft. The morphogen possibility has previously been supported by evidence that graded distributions of exogenous retinoid are the most effective at inducing pattern duplication¹¹, but this could reflect the time of exposure of neighbouring tissue that is induced to become a ZPA (see page 1919 of ref. 12). The evidence that RA and ddRA are extractable from the limb bud, and that the former at least is present at somewhat higher levels in posterior fragments¹³, raises questions about the functional significance of endogenous retinoids, and whether they are responsible for inducing the ZPA *in vivo*.

The other key system for evaluating the effects of retinoic acid on limb morphogenesis is regeneration in urodele amphibians. Amputation of the limb at any level on the proximodistal (PD) axis from shoulder to finger-tip is followed by formation of blastemal (progenitor) tissue at the plane of amputation. Blastemal cells normally give rise only to structures distal to their point of origin, but RA is able to respecify them to more proximal values. Thus a wrist blastema, which normally forms a hand, may regenerate an entire arm¹⁴. These remarkable effects on the PD axis have recently been extended by Stocum and colleagues¹⁵ to include unidirectional alterations of positional value on the transverse axes in certain experimental contexts. Retinoic acid is therefore able to affect all three axes.

Histological evidence

In view of that, and the lack of evidence for a signalling centre akin to the ZPA, along with the absence of data on endogenous retinoid levels, the case for RA as an endogenous morphogen in limb regeneration has not found strong advocates². The histological evidence suggests that RA treatment of distal regenerates promotes changes in the tissue morphology of the blastema¹⁶, and it is possible that the effects of RA on axial specification are secondary to its ability to influence the extent of the blastema.

In this sense, retinoic acid could be said to establish boundaries, just as local induction of a polarizing region sets the posterior boundary in the chick limb. But what are the molecular bases of these effects, how does the ZPA work and how is position encoded in the blastema and the avian bud? Discussion of these issues has been greatly stimulated by the work on RA, but at present its election to vanguard membership of the club of vertebrate morphogens can be questioned. □

Jeremy Brockes is in the Ludwig Institute for Cancer Research, Middlesex Hospital/University College London Branch, 91 Riding House Street, London W1P 8BT, UK.

1. Eichele, G. *Trends Genet.* **5**, 246–251 (1989).
2. Brockes, J. P. *Neuron* **2**, 1285–1294, (1989).
3. Brockes, J. P. *Nature* **345**, 766–768 (1990).
4. Wanek, N., Gardiner, D.M., Muneoka, K. & Bryant, S.V. *Nature* **350**, 81–83 (1991).
5. Noji, S. *et al.* *Nature* **350**, 83–86 (1991).
6. Summerbell, D. & Harvey, F. *Prog. clin. biol. Res.* **110**, 109–118 (1983).
7. Smith, S. M. & Eichele, G. *Development* **111**, 245–252 (1991).
8. Rowe, A., Richman, J.M. & Brickell, P. *Development* (in the press).
9. de The, H. *et al.* *Nature* **343**, 177–180 (1990).
10. Sucov, H.M., Murakami, K.K. & Evans, R.M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5392–5396 (1990).
11. Thaller, C. & Eichele, G. *Nature* **345**, 815–819 (1990).
12. Eichele, G., Tickle, C. & Alberts, B.M. *J. Cell Biol.* **101**, 1913–1920 (1985).
13. Thaller, C. & Eichele, G. *Nature* **327**, 625–628 (1987).
14. Maden, M. *Nature* **295**, 672–675 (1982).
15. Ludolph, D.C., Cameron, J.A. & Stocum, D.L. *Dev. Biol.* **140**, 41–52 (1990).
16. Kim, W.-S. & Stocum, D.L. *Wilhelm Roux Arch. dev. Biol.* **195**, 243–251 (1986).

Turning on and turning off the sense of smell

Brian Burchell

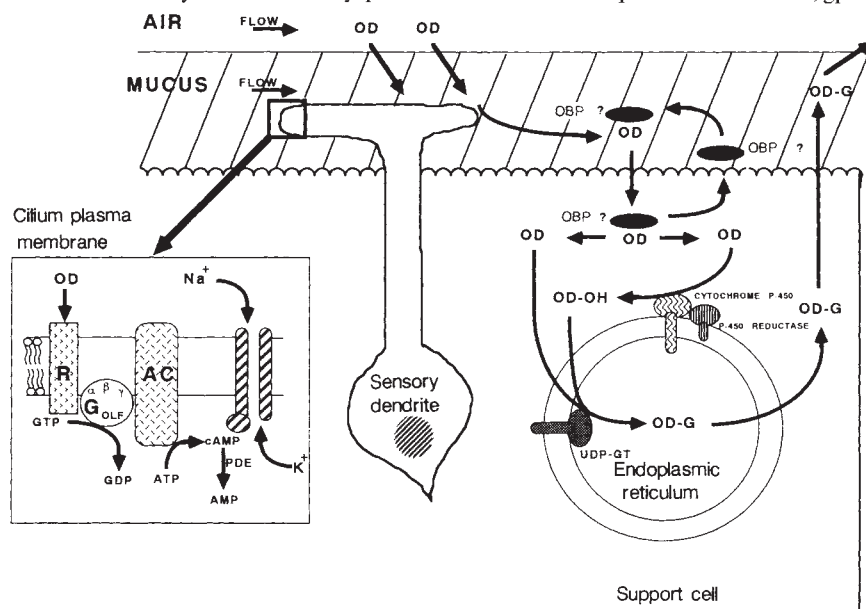
THE mammalian olfactory system is a highly specialized molecular recognition unit which exhibits exquisite discrimination and sensitivity. More than ten thousand distinct odours of low-molecular-mass volatile compounds can be detected with high fidelity, small changes of chemical structure producing patterns of neuronal activity that are relayed to the brain and decoded as distinct odours. Primary signal transduction occurs in the olfactory neuroepithelium, while the processing of sensory information occurs in the olfactory bulb^{1,2}. Most recent progress has been concerned with the molecular recognition of odours, and the subsequent signal generation and termination, as illustrated by a paper by Bakalyar and Reed in *Science*³ and one from Lancet's group in *Nature*⁴. Indeed, the past 12 months has been a vintage period for the molecular cloning of specific olfactory transduction proteins, with reports from two laboratories of the isolation of a cation channel gated by cyclic AMP^{5,6}.

Rainbow trout

Neuroepithelial cilia, believed to be the site of transmembrane signalling, were isolated from the olfactory rosette of rainbow trout ten years ago¹. The chemical recognition systems in these membranes are thought to be quite diverse, suggesting that there may be several receptor proteins which operate through common transducing components. The identification of odorant-stimulated, G protein-dependent adenylatecyclase in isolated olfactory cilia, at more than ten times the level in brain membranes^{7,8}, subsequently provided the experimental basis for the most recent work. Almost all odorants increase adenylatecyclase activity and several regulatory G proteins have been localized to olfactory cilia. Pfeuffer *et al.*⁹ purified a novel (type III) adenylatecyclase from rat olfactory cilia which, when cloned by Bakalyar and Reed, was demonstrated to be present only in olfactory cilia. Type III adenylatecyclase co-localizes with the olfactory specific G protein, G_{OLF} . Although their interaction has not been demonstrated, G_s -proteins can activate type III adenylatecyclase and regulate the response of cyclic AMP — high levels of type III adenylatecyclase in olfactory cilia rapidly generate high levels of cAMP for neuronal activity in response to exogenous odorant¹⁰.

Olfactory cilia and visual rod photoreceptors contain cation-selective ion channels which are directly and cooperatively opened by cyclic nucleotides¹¹. Complementary DNAs encoding ion-channel proteins from bovine and rat olfactory epithelium have been cloned by Ludwig *et al.*⁵ using a

probe for the cyclic nucleotide-binding site of the rod gated channel and by Dhallan *et al.*⁶ with low-stringency screening using the rod probe. Dhallan *et al.* went on to examine the electrophysiological properties of the rat olfactory ion channel by patch-



Model for the transduction and termination of olfactory signals. Odorants (OD) in the air are mostly lipid soluble, enter the mucus flow and reach neuroepithelial cilia. Odorant interaction with a receptor (R) stimulates adenylatecyclase (AC) by interaction with G proteins (G_{OLF}) to produce cAMP which binds with the cation channel and promotes neural signalling (see inset). Alternatively odorants may bind to olfactory binding proteins (OBP) for transmission to receptor(s) or for detoxification. Odorant signals are terminated by hydroxylation and glucuronidation in the endoplasmic reticulum of the neuroepithelial support cells. Inactive odorant glucuronides (OD-G) are released from the cells to the mucus flow or vasculature. PDE, cAMP phosphodiesterase; UDP-GT, UDP-glucuronosyltransferase.

clamping human embryonic kidney cells transiently expressing the transfected cDNA. A mixed (Na^+/K^+) cation-conducting channel was obtained. Thus neuronal signals can be transmitted to the information-processing unit in the olfactory bulb.

Odorant sensing is generally a rapid but transient process, requiring instant response and also speedy termination of the signal. How is the odorant response terminated? A simple mechanism would be to remove excess odorant by bioactivation and accelerated excretion of the modified chemical. Specific examples of xenobiotic detoxicating systems, the monooxygenases, which have cytochrome P450 as the central catalytic unit, were identified in olfactory tissue some years ago¹², and were subsequently purified and further characterized by cloning of the cDNA¹³. However, the two cytochrome P450s, members of an extensive gene family identified from studies of other tissues, are well known as bioactivators of inert chemicals producing substrates for xenobiotic conjugating enzymes. The odorant might be

modified by monooxygenases, but remain an active chemical stimulant.

A significant discovery of 1990 was the identification of an olfactory UDP-glucuronosyltransferase that catalyses odorant glucuronidation to inactive chemicals. Lancet's group identified two major glycoproteins, gp56 and gp52, in bovine olfactory deciliated epithelium membrane¹⁴ — sequences derived from peptides isolated from gp56 are very similar to those of the multigene family of UDP-glucuronosyltransferases described in the sequence database^{14,15}; gp52 is

similar to the cytochrome P450 IIA family. Both proteins are localized in the endoplasmic reticulum of olfactory epithelium. The subsequent cloning and expression of the olfactory UDP-glucuronosyltransferase, reported by the same group last week⁴, reveals that this transferase preferentially catalyses the glucuronidation of a range of odorants. Glucuronidation also prevents odorant-stimulated increase in adenylatecyclase activity, indicating that olfactory odorant UDP-glucuronosyltransferase provides an effective mechanism of signal termination. But odorants would have to reach the endoplasmic reticulum in olfactory epithelium to be effectively bioinactivated.

Odorant receptors

How do the diverse odorants reach different membrane surfaces to initiate a response and then be subsequently inactivated? Key receptor/transporter proteins have yet to be identified. Odorant-binding protein (OBP), characterized by Snyder and colleagues² and localized to Steno's gland, may carry odor-

ants to the receptor surface or alternatively may 'scavenge' excess odorant for transport to endoplasmic reticulum for biotransformation, similar to the textbook role of ligandin in the transport of bilirubin. Lancet's team has proposed that a transmembrane glycoprotein of relative molecular mass 95,000, which is enriched in frog olfactory epithelium, has properties consistent with those of a membrane receptor, but has so far been unable to demonstrate that it interacts directly with odorants. Genetic studies may aid the identification of a number of distinct

odorant receptors — several dozen separate genetically determined anosmias (conditions resulting in loss of the sense of smell) have been reported so far, and receptor diversity may therefore run into the thousands. It could be, though, that odorant interaction with a more limited array of different receptors provides the subtlety of response required. □

Brian Burchell is in the University Department of Biochemical Medicine, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK.

experimental control achieved by Lockner *et al.*, and their ability to trace accurately the evolution of the granite's fracture, allowed the authors to show that shear localization started only after the rock had reached its peak strength and had already started to soften homogeneously. This is in agreement with the behaviour predicted by Rudnicki and Rice's 'bifurcation' analysis, and in disagreement with experiments suggesting that shear localization starts while the material is still hardening.

Most continuum descriptions do not have any prescribed length scales, so that bifurcation analysis cannot be used to predict the geometry of the shear band. As the width of the fault zone in an experimentally deformed sample is clearly smaller than that of a geological fault, it is important to have a theoretical framework to translate the laboratory results into tectonically interesting predictions. The data obtained by Lockner *et al.*¹ can provide important constraints on the formulation of more elaborate continuum models for brittle rock deformation.

Following microstructural observations that stress-induced microcracking occurs stably and that shear localization involves the coalescence of numerous microcracks, models were developed^{6,7} treating fracture mechanics at the scale of individual grains. The theoretical treatment of crack interaction and coalescence is particularly difficult, and microstructural observations on different rock types (such as granite, marble and sandstone) indicate that the localization process is sensitive to porosity, mineralogy and pressure. Lockner *et al.*⁸ have reported important differences in the spatial distribution of acoustic-emission activities between granite and sandstone.

With improved resolution in the laboratory, a question naturally arises as to the relation of the laboratory results to crustal scale faulting. McGarr *et al.*⁹ observed localized deformation in South Africa, due to mining-induced earthquakes, which is very similar to that mapped by Lockner *et al.* And field maps of shear bands in the Navajo and

ROCK MECHANICS

Action replay for fracture

Teng-fong Wong

BRITTLE fracture can be slowed down so greatly, report Lockner *et al.* on page 39 of this issue¹, that rather than taking a mere fraction of a second, the process can be drawn out over hours. This enabled the authors to study the way faults nucleate and grow in stressed granite, making the experiment a miniature imitation of the process of earthquakes.

Put a rock under uniform pressure and it will generally deform uniformly until some critical stress is reached, at which it will yield locally along a single plane. This shear localization occurs on all scales, from small rock samples studied in the laboratory up to whole crustal fault zones, the sources of earthquakes; as well as arising in rocks, it is found in metals and ceramics². Tectonic deformation of the Earth's crustal plates occurs over a broad spectrum of temperature, pressure and timescale. At one end of the spectrum, shear localization in brittle rocks under upper-crustal conditions of pressure and temperature results in the nucleation and propagation of geological faults. Motivated by observations on the systematic relation between fault orientation in rocks and the relative magnitude of the principal stresses, Anderson³ formulated his theory of faulting which has provided a physical basis for the classification of faulting styles into normal, thrust and strike-slip respectively. Although there are abundant data on the peak stress levels at the onset of faulting, little is known about the detailed evolution of the shear localization process that culminates in a throughgoing fault zone. In their new experiment, Lockner *et al.* combine sophisticated nondestructive testing, source location, and servo-control techniques,

originally developed for materials science and seismology, to map out the spatial and temporal development of faulting in granite samples.

Laboratory experiments of this kind usually simulate crustal stress conditions by subjecting the sample to an overall compressive loading in all three directions. In the axial loading configuration used by Lockner *et al.*, the sample's load-supporting capability typically reaches a maximum near the onset of shear localization. As shear localization develops, there is a concomitant degradation of the rock's overall strength, and the faulting process becomes catastrophic unless the surplus of elastic energy stored in the test machine is quickly relieved.

One way of stabilizing this post-failure deformation is to construct an anomalously stiff loading frame so that it stores little elastic energy. Researchers at Kyoto University⁴ showed that an alternative approach, using feedback control using fast electronics, can stabilize the rock failure. They used volumetric strain analysis and acoustic emission from the rock to monitor the faulting. Lockner *et al.* adopt a similar approach, recording the acoustic emission rate both for feedback control to slow the fracturing process and for a form of three-dimensional source location to plot the progress of the fault's hypocentres.

Regarding rocks as continuous media, the onset of shear localization can be viewed as an instability in the description of homogeneous deformation. In a seminal study, Rudnicki and Rice⁵ showed that the onset of shear localization can be related to the onset of nonunique solutions in the homogeneous stress-strain relations. Significantly, the ex-

1. Lockner, D.A., Byerlee, J.D., Kuksenko, V., Ponomarev, A. & Sidorin, A. *Nature* **350**, 39–42 (1991).
2. National Materials Advisory Board *Fracture in Compression of Brittle Solids* Report NBAB-404 (National Academy, Washington DC, 1983).
3. Anderson, E.M. *The Dynamics of Faulting* (Oliver & Boyd, Edinburgh, 1951).
4. Terada, M., Yanagidani, T. & Ehara, S. in *3rd Conf. on Acoustic Emission/Microseismic Activity in Geologic Structures and Materials*, (eds Hardy, H.R. & Leighton, F.W.) 159–171 (Trans-Technical, Causthal-Zellerfeld, 1984).
5. Rudnicki, J.W. & Rice, J.R. *J. Mech. Phys. Solid* **23**, 371–395 (1975).
6. Horii, H. & Nemat-Nasser, S. *J. geophys. Res.* **90**, 3105–3125 (1985).
7. Ashby, M.F. & Hallam, S.D. *Acta metall.* **34**, 497–510 (1986).
8. Lockner, D.A., Byerlee, J.D., Kuksenko, V., Ponomarev, A. & Sidorin, A. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-f.) (Academic, in the press).
9. McGarr, A., Spottiswoode, S.M., Gay, N.G. & Ortlepp, W.D. *J. geophys. Res.* **84**, 2251–2261 (1979).
10. Aydin, A. & Johnson, A.M. *J. struct. Geol.* **5**, 19–31 (1983).
11. Segall, P. & Pollard, D.D. *J. geophys. Res.* **88**, 555–568 (1983).

Entrada sandstones in Utah¹⁰ can be adequately interpreted by Rudnicki and Rice's model. However, it should be noted that the boundary conditions for tectonic processes can be extremely complex. For example¹¹, a granodiorite fault zone in the central Sierra Nevada probably nucleated on a preexisting weak zone comprising extensional joints, and shear localization evolved with the rotation of the tectonic stress field over time.

The shear fracture energy is an important parameter in the fracture mechanics of the seismogenic process. The technique developed by Lockner *et al.* has a lot of potential. More extensive studies on different rock

types should provide important insights on the scaling of the fracture mechanics parameter to earthquakes. The main uncertainty at this point is the physical basis for the empirical calibration of energy release in terms of the acoustic emission amplitude. The resolution of this question hinges on a fundamental understanding of the source dynamics of acoustic emission, a topic being intensively pursued in materials science. □

Teng-fong Wong is in the Department of Earth and Space Sciences, State University of New York, Stony Brook, New York 11794, USA.

CIRCADIAN RHYTHMS

Resetting the human clock

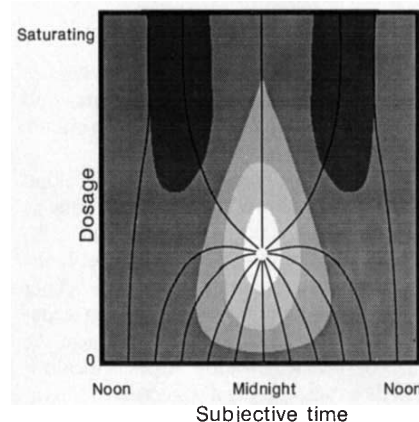
A. T. Winfree

THE experiments of Jewett *et al.*¹, reported on page 59 of this issue, bring to bear on human medicine a third of a century of theory and experiments characterizing the dynamics of circadian clocks. The authors have now verified in humans a hypothesis that was advanced in the late 1960s, and confirmed experimentally in plants and many species of invertebrates during the 1970s. As well as providing a key to the management of jet-lag, and presumably of other sleep disorders, this episode is unusual in the history of biomedical sciences in that theory has consistently led the way.

The 1960 Cold Spring Harbor Symposium on Quantitative Biology was devoted to circadian clocks. H. Kalmus and L. A. Wigglesworth there suggested that nonlinear dynamics might usefully organize inquiry in the subject more quickly than biochemistry, and they proposed the 'attracting limit cycle oscillator' as a close metaphor for clock dynamics. Although little evidence ever accrued in support of the limit cycle, a related but more abstract model based on nonlinear dynamics was later put forward and proved to predict the essence of clock behaviour. According to this model, the internal state of the clock may be understood as a point moving smoothly along a closed-cycle trajectory in a euclidean state-space of two or more dimensions. All circadian clocks have evolved mechanisms by which they are frequently reset to keep pace with changes in the local environment, most notably by being sensitive to the intensity of light. In the model, the effect of light is interpreted as a nonlinear bending of the free-running trajectory. Topological inference from such simple principles leads to surprising but testable implications, which are summarized in the figure.

First, there must be a unique instant in the circadian cycle during which a singular duration of exposure to light, instead of resetting the clock, terminates all rhythmicity. Second, that singular combination of timing and magnitude may vary slightly among individ-

uals, and slight departures from it will reset the clock by large and arbitrary amounts. Third, a striking depression of clock amplitude should accompany these events. These implications had never before been observed, however, and seemed implausible.



Amplitude and phase resetting in a generic circadian clock. Vertical axis: ranging from 0 to a saturating dose, the magnitude of a resetting stimulus. Horizontal axis: spanning one circadian clock cycle, the subject's internal circadian phase at the moment when the resetting stimulus begins. Shading: new circadian clock amplitude after that stimulus (1 along the horizontal axis, 0 at the singularity). The new phase to which the circadian clock is reset by the end of the stimulus is discovered by tracing the nearest contour line back to the horizontal axis.

Alternative, more widely accepted, notions of clock dynamics led to very different expectations. For example, the 'biological clocks' governing the menstrual cycle and the timing of nuclear and cell division have quite different dynamics and patterns of response to resetting stimuli. Nevertheless, the predictions were eventually borne out in experiments with the fruitfly *Drosophila* and later with other species. But it remained possible to doubt their pertinence to human medicine.

Then it was observed that the peculiar mathematical equation originally used to describe phase resetting in the fruitfly fitted

human data equally well². Does the amplitude response in humans also conform to the generic pattern? Jewett *et al.*¹ engaged the cooperation of 14 human subjects to determine whether the amplitude of the circadian rhythm is labile and falls off towards zero as the timing and magnitude of the stimulus approach the singular combination. Previously, such amplitude behaviour had been investigated (and confirmed) only in a flower and in two species of dipteran insect. The careful experimentation with humans by Jewett *et al.*¹ and by Czeisler *et al.*² now shows that the three decades of work with non-human species has indeed been relevant to us.

Humans differ from laboratory organisms in that only seldom and irregularly do they expose themselves to the brightness of light required by their circadian clocks on a regular schedule for reliable resetting³. Such irregularity, exacerbated by abrupt travel across time zones, implies that such stimuli that we do indeed receive may correspond to sampling of the inner regions shown in the figure: that is, of times around subjective midnight and dosages at way below saturation, where phase resetting is large and is accompanied by depression of amplitude. Given a version of this generic diagram calibrated for humans, we would know exactly how circadian phase and amplitude are affected by any stimulus of given timing and magnitude. In principle this provides the indispensable key to the rescheduling that may be required for control of jet-lag, shift work and sleep disorders.

The experiment of Jewett *et al.*¹, together with that of Czeisler *et al.*², go a long way towards providing this calibration. They place the singularity, the organizing centre of the pattern, around subjective midnight, as is usual in diurnal species. And they show that the singular dose of light (10-hour exposure at intensities comparable to those at dawn or dusk) is about 50 times greater than that previously hinted at from experiments with other vertebrate species^{4,5}. If stimuli of this magnitude are needed for substantial resetting of the clock, a more convenient dosage for management of jet-lag and shift work, for example, would be briefer and at higher intensity; my own habit for 10–15 years has been to take about two hours of full sunlight at a time crudely estimated to provide the phase shift necessary. The new work means it may now be possible to make a more quantitative and accurate prescription. □

A. T. Winfree is in the Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85718, USA.

1. Jewett, M.E., Kronauer, R.E. & Czeisler, C.A. *Nature* **350**, 59–62 (1991).
2. Czeisler, C.A. *et al. Science* **244**, 1328–1333 (1989).
3. Savides, T.J., Messin, S., Senger, C. & Kripke, D.F. *Physiol. Behav.* **38**, 571–574 (1986).
4. Gander, P.H. & Lewis, R.D. *Am. J. Physiol.* **245**, R10–17 (1983).
5. Eskin, A. in *Biochronometry* (ed. Menaker, M.) 55–80 (National Academy of Sciences, Washington DC, 1971).

Iron determinations

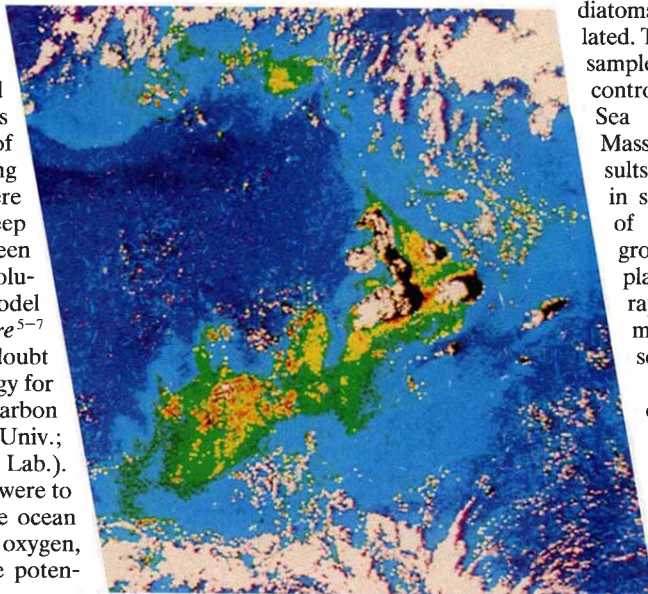
Philippa Lloyd

OCEANOGRAPHERS have long been puzzled by areas of the ocean where there are high levels of nutrients (nitrate and phosphate), and yet the phytoplankton biomass is surprisingly low. Poor light, low temperature, zooplankton grazing and trace-metal limitation (or toxicity) have been blamed for the low productivity, but none convincingly. J. Martin and colleagues (Moss Landing Lab.) have evidence¹⁻⁴ indicating that photosynthesis is limited by a shortage of iron in these areas, and suggest that fertilizing the oceans with iron may stimulate the growth of giant algal blooms, thereby removing carbon dioxide from the atmosphere and sequestering carbon in the deep ocean. Such an approach has been hailed as a possible technological solution to the greenhouse effect. But model calculations presented both in *Nature*⁵⁻⁷ and at a meeting* last month, cast doubt on the effectiveness of such a strategy for reducing atmospheric levels of carbon dioxide (J. Sarmiento, Princeton Univ.; T.-H. Peng, Oak Ridge National Lab.). Indeed, if the fertilization proposal were to be carried out, certain areas of the ocean might become severely depleted in oxygen, models indicate, which could have potentially catastrophic effects.

Iron is important for algal growth, but only trace amounts are needed, which might imply that the phytoplankton's iron requirements could easily be met. But in the oxygenated sea water, soluble Fe(II) is oxidized to insoluble Fe(III). Supply of iron is not a problem in coastal areas, mainly because of resuspension of iron-rich sediments on the continental shelf. In offshore regions, iron is chiefly supplied by windborne dust containing iron (as well as other trace metals) originating from arid areas in Asia. A significant fraction of the iron in atmospheric dust is believed to be soluble Fe(II) and therefore readily available for uptake by phytoplankton. High fluxes of iron-rich dust are found off the coast of Asia, China, West Africa, Australia and the Caribbean, but over the Southern Ocean, the central equatorial Pacific and the northern Pacific, fluxes are low. Intriguingly, these are the upwelling regions of the ocean where the conditions of high nutrients and low productivity are observed (R. Duce, Univ. Rhode Island). This bolsters Martin's argument.

Measurements of the carbon dioxide concentration in ancient air trapped in polar ice show that CO₂ levels were lower during glacial periods; moreover, the supply of iron-rich atmospheric dust was then about 50 times higher than it is now. Martin⁸ suggests that the increased availability of iron during

glacial periods significantly enhanced phytoplankton growth, thereby contributing to the removal of atmospheric CO₂. But productivity seems not to have been enhanced in the nutrient-rich Southern Ocean (W. H. Berger, Scripps Inst. of Oceanography). Nonetheless, the Antarctic seems to have been productive only for the past two million



False-colour satellite image showing algal blooms around the Galapagos — "a natural iron-fertilization experiment" according to Martin, as plumes to the southwest (orange-green) may be stimulated by iron-enriched currents. (White, clouds; blue, unproductive ocean.) Courtesy Gene Feldman.

years or so, and there have been significant amounts of dust in the atmosphere for only the past 6 million years. So possibly there is a role for iron in stimulating productivity, although not quite as envisaged by Martin. As an alternative to iron limitation, J. J. Cullen (Bigelow Lab. for Ocean Sciences) considered whether upwelling may cause poor productivity: phytoplankton might be unable to adapt sufficiently quickly to their changing light environment in rapidly upwelling regions such as the central Pacific. But field studies show they can: the growth rate and chemical composition of the phytoplankton, estimated from changes in chlorophyll during incubation of samples, were found to be high (just < 1 division per day). The availability of trace metals did not limit the growth of the dominant phytoplankton species, and Cullen concluded that grazing by microzooplankton regulated the standing crop. That the availability of iron might control the growth of certain larger species, and hence the food web structure, could not, however, be ruled out.

Results of a study in the sub-Arctic Pacific show a marked lack of variation in the standing stock of phytoplankton, despite large changes in the solar radiation received, the depth of the mixed layer and temperature.

Here too the specific growth rates of the phytoplankton were found to be high. A model invoking not only a balance between growth and grazing, but also the preferential uptake of ammonium (regenerated production), as opposed to nitrate uptake (new production), can explain the observations without direct recourse to iron limitation (B. Frost, Univ. Washington).

In contrast, in experiments in which iron was added to bottled samples of ocean water, nitrate was taken up more efficiently than in control samples and the growth of large diatoms, in particular, seemed to be stimulated. The species composition of the treated samples was very different from that of the controls (A. Buma, Netherlands Inst. for Sea Research; Martin; N. Price, Massachusetts Inst. Technology). Such results lead to the hypothesis that in some areas of the ocean, low levels of dissolved iron might inhibit the growth of large nitrate-utilizing phytoplankton and instead promote a rapidly cycling ammonium-fuelled microbial loop (F. Morel, Massachusetts Inst. Technology).

The interpretation of these bottle experiments, however, is controversial: iron may indeed be limiting; or the addition of iron could have a negative effect on grazers and the consequent relaxation in grazing pressure could yield the observed increase in chlorophyll; or iron could be limiting only for the larger bloom-forming phytoplankton (as suggested above); or it could be that iron addition reduces the lag between nutrient uptake and population growth. In addition, the sampling procedure for bottle experiments leads to the exclusion of large grazers and hence the community structure is different from that in the open ocean in both the control and treated samples. A more profitable approach (P. Falkowski, Brookhaven National Lab.) is to examine the cells of various oceanic phytoplankton species and look for appropriate markers of iron limitation.

Although iron fertilization does not seem to be even a partial cure for the greenhouse effect, the debate over the role of iron in regulating productivity demands considerable further work, with perhaps small-scale iron fertilization experiments in the ocean, as well as studies to examine the chemistry of iron in sea water and the physiology of the relevant phytoplankton species. □

Philippa Lloyd is an assistant editor of *Nature*.

*American Society of Limnology and Oceanography special symposium, San Marcos, California, 22–24 February 1991.

1. Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
2. Martin, J. H. & Gordon, R. M. *Deep-Sea Res.* **36**, 649–680 (1989).
3. Martin, J. H. *et al.* *Nature* **345**, 156–158 (1990).
4. Davies, A. G. *Nature* **345**, 114–115 (1990).
5. Peng, T.-H. & Broecker, W. S. *Nature* **349**, 227–229 (1991).
6. Keir, R. S. *Nature* **349**, 198 (1991).
7. Joos, F., Sarmiento, J. & Siegenthaler, U. *Nature* **349**, 772–775 (1991).
8. Martin, J. H. *Paleoceanography* **5**, 1–13 (1990).

Aromaticity revisited

Patrick Fowler

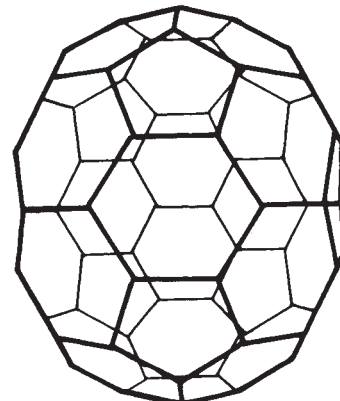
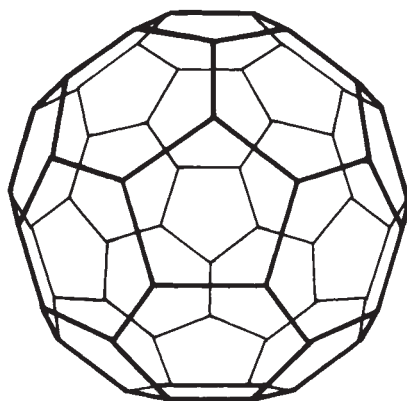
ALLOTROPES of the elements are not discovered every day, and the new synthetic fullerites¹ which appear to consist of close-packed pure-carbon footballs are attracting attention from experimentalists and theoreticians alike. Measurements of all kinds on these solids are being published already and teams at AT&T Bell Laboratories and Cornell University report one more on page 46 of this issue²: the magnetic susceptibility of C_{60} and C_{70} powders. This particular measurement was eagerly awaited because of the debate on the degree of electron delocalization — or aromaticity — of buckminsterfullerene. In the event, C_{60} is found to be diamagnetic, but less so than expected on the crude 'ring-current' model, whereas C_{70} is strongly diamagnetic and conforms more closely to the naive picture of charge moving freely through the bonds. Perhaps not surprisingly, Haddon and co-workers² conclude that it may be difficult to fit the fullerenes into traditional aromatic or non-aromatic categories.

It is only recently that attention has shifted from the structure of C_{60} to its properties. The C_{60} molecule was first identified in the products of laser-ablated graphite³. The evidence was a dominant, long-lived peak in the mass spectrum at mass 720, which was taken to be the signature of an especially stable shell of 60 carbon atoms making up the 12 pentagonal and 20 hexagonal faces of a truncated icosahedron (see figure). Theoreticians seized on this elegant structure, a finite version of the graphite sheet, and made the obvious connection with planar conjugated systems. It became a commonplace that this was an aromatic molecule and a three-dimensional analogue of benzene. The large number of hexagons would preserve much of the increased binding energy gained in graphite through electron delocalization and simple empirical (Hückel) calculations predicted a structure with 60 long (roughly single) and 30 short (roughly double) carbon-carbon bonds, corresponding to a sheath of π -electrons enclosing the edges of the polyhedron.

More sophisticated *ab initio* and semi-empirical electronic structure calculations added detail to the picture. We now have good estimates of the bond lengths⁴ and confirmation that C_{60} is more strongly bound than other all-carbon molecules. The broad features of the electronic structure, the electronic configuration, the sequence of ionization energies and symmetry-based selection rules for ultraviolet transitions are already present in the Hückel picture. This crude topological approach also accounts for the fact that C_{60} is not alone. An infinite number of triply-connected hexagon-plus-pentagon cages can be imagined. These fullerenes⁵ explain the forest of minor mass peaks found

in the laser-ablation experiment.

Just as conjugated hydrocarbons show special stability for 'magic' numbers, the electrons of pseudospherical carbon clusters fall neatly into closed shells for certain vertex numbers and molecular symmetries⁶. Icosahedral clusters with 180, 240, 420, 540 atoms, and so on, cylindrical clusters with 10 or 12 times $(7+3k)$ atoms share the magic property of C_{60} . At least one stable cluster can be imagined for all 'leapfrog'⁶ numbers $(60+6k)$ other than 66, paralleling the famous Hückel $4n+2$ rule for planar hydrocarbon rings. The egg- (or rugby-ball) shaped C_{70} is an example of the cylinder class (see figure); it remains to be seen whether



The principal members of the fullerene family: soccerball-shaped C_{60} and egg-shaped C_{70} .

the minor C_{84} component of fullerite is cylindrical or tetrahedral in shape.

The problem facing the investigator of C_{60} is an odd one; like a certain fictional hero this molecule is (almost) without qualities. Most of the observables used to fingerprint molecules are based on anisotropy: a molecule may have a dipole moment and so give a microwave spectrum; inequivalence amongst its atoms may generate informative NMR splitting patterns; or it may have active sites that react with probe molecules. C_{60} yields none of these clues. It could be hardly more isotropic (apart from the hypothetical 120-atom great rhombicosidodecahedron, no larger three-dimensional cluster composed entirely of equivalent atoms is possible) and it is inert under chemical attack. C_{60} has to be diagnosed by the simplicity of its few spectral fingerprints.

Crucial tests of the structural hypothesis awaited the synthesis of macroscopic amounts. Fullerite can now be made in gram quantities by a relatively simple technique using an arc struck between graphite electrodes¹. The crystals retain the distinctive mass spectrum. Definitive X-ray structural characterization is made difficult by disorder, but infrared¹, Raman⁷ and NMR⁸ spectra have been recorded. Icosahedral C_{60} should have only four infrared-active fun-

damentals; four are found in thin-film, gas-phase and matrix experiments. Of ten possible Raman-active modes, three have been observed. Perhaps most convincingly of all, the ^{13}C NMR solution spectrum shows only the single peak expected for a molecule made up of equivalent atoms. C_{70} shows a richer infrared spectrum and five NMR signals corresponding to the five distinct carbon sites in the postulated structure⁹.

The NMR experiments reveal some interesting contrasts with the susceptibility measurements reported² in this issue. Haddon *et al.* confirm that both C_{60} and C_{70} are diamagnetic: the induced circulation in the electron cloud opposes the external field. If π electrons move freely then large 'ring currents' are expected, otherwise there will be low net magnetism. The physical reality of ring currents is debatable, but bond-susceptibility models can also rationalize exaltation

of diamagnetism in aromatics. Haddon *et al.*² use a ring-current semi-empirical model which is in good agreement with their experiment to predict small currents in C_{60} (less than in benzene) and large ones in C_{70} . On this model, C_{70} appears aromatic but C_{60} does not.

On the other hand the NMR carbon signal of C_{60} lies 143 parts per million (p.p.m.) downfield of TMS, the reference standard for NMR, and the five peaks for C_{70} occur between 130 and 150 p.p.m. So the magnetic environment of an average carbon in C_{70} is very similar to that in C_{60} . Second, the C_{60} signal is not far (downfield by 14 p.p.m.) from the prototypical aromatic molecule — benzene. On an absolute scale (with respect to the bare nuclei) carbon in benzene is shielded from the NMR magnetic field by 57 p.p.m. by the delocalized electron cloud, and carbon in C_{60} is diamagnetically shielded by 43 p.p.m. This value agrees with one that can be extrapolated from *ad initio* Hartree-Fock calculations⁹ on C_{60} (40 p.p.m.). There seems then to be a conflict between the similarity in NMR terms of C_{60} and C_{70} and the difference in their magnetic susceptibility. The two properties are not mathematically linked, but both reflect the response of an electron cloud to an applied magnetic field.

Another probe of aromatic character in

NMR is the anisotropy in the carbon chemical shift; although C_{60} is isotropic, the sixty sites have individually only mirror symmetry and therefore has three principal components of chemical shift. At room temperature these are averaged by rotation, but they can be extracted by fitting line envelopes in low-temperature ^{13}C NMR studies on powder samples¹⁰. Components of 220, 186 and 40 p.p.m. are found. This 'axial' pattern and large splitting are characteristic of aromatic carbons.

The question of the aromaticity of the fullerenes is clouded by the fact that we have two different pictures in mind; a fullerene is like graphite in that it has hexagonal rings and fits a delocalized molecular-orbital description; on the other hand it is like benzene because it has a closed electronic shell and a substantial 'band gap' between filled and empty levels. The two are ultimately incompatible⁷ — as the cage size increases the delocalization stabilization rises but the band gap

must fall, in line with the semi-metallic nature of graphite. Clearly the main point is whether either analogy will lead to prediction or rationalization of observables. To judge from the magnetic properties measured so far, sometimes they will and sometimes they won't. □

Patrick Fowler is in the Department of Chemistry, University of Exeter, Stocker Road, Exeter EX4 4QD, UK.

1. Krätschmer, W., Lamb, L.D., Fostiropoulos, K. & Huffman, D.R. *Nature* **347**, 354–358 (1990).
2. Haddon, R.C. *et al.* *Nature* **350**, 46–47 (1991).
3. Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F. & Smalley, R.E. *Nature* **318**, 162–163 (1985).
4. Scuseria, G.E. *Chem. Phys. Lett.* **176**, 423–427 (1991).
5. Kroto, H.W. *Nature* **320**, 529–531 (1987).
6. Fowler, P.W. *Chem. Phys. Lett.* **131**, 444–449 (1986).
7. Bethune, D.S., Meijer, G., Tang, W.C., & Rosen, H.J. *Chem. Phys. Lett.* **174**, 219–222 (1990).
8. Taylor, R., Hare, J.P., Abdul-Sada, A.K., & Kroto, H.W. *J. Chem. Soc. Chem. Comm.* 1423–1425 (1990).
9. Fowler, P.W., Lazzeretti, P. & Zanasi, R. *Chem. Phys. Lett.* **165**, 79–86 (1990).
10. Yannoni, C.S., Johnson, R.D., Meijer, G., Bethune, D.S. & Salem, J.R. *J. Phys. Chem.* **95**, 9–10 (1991).

HIGH-TEMPERATURE SUPERCONDUCTORS

Light at the end of the tunnel

M. R. Beasley

ONE of the ideas to be reinvigorated by the discovery of high-temperature superconductors a few years ago was the possibility of basing computer electronics on superconducting devices — particularly the 'tunnel junction'. As with previous attempts with conventional superconductors, turning the new high-temperature materials into tunnel junctions has proved hard, leading some to argue that the sensitivity of the tunnelling process to surface conditions will make it impossible. Nevertheless, three groups have now reported^{1–3} substantial progress, so that although the prognosis is still ambivalent, the fog is partially clearing.

Why are the tunnel junctions so interesting? They are the device of choice in superconductive electronics, at least as practiced with the low-temperature superconductors such as niobium. Also, the details of the electrical characteristics of tunnel junctions yield information regarding the nature of the electronic interaction giving rise to the superconductivity — and hence the mechanism of the superconductivity itself.

To make a tunnel junction, in which the electrons pass from one conducting electrode to another by quantum mechanically tunnelling through a classically impassible region, one has to separate two layers of superconductor by a very thin tunnel barrier — a layer of insulating material. The problem has been forming such a barrier without affecting the underlying superconducting material. It was known early on that thin layers of gold or silver have a minimal chemical effect on the high-temperature, oxide superconductors' surfaces. So, many groups have tried to build their insulating tunnel barrier on such thin normal-metal overlayers

and study the underlying high-temperature superconductor through its influence on the overlayer.

Superconductivity is induced in the nominally nonsuperconducting metal overlayer through the phenomenon of the proximity effect: superconducting electrons penetrate into the overlayer to endow it with superconductivity. It is this induced superconductivity that is actually observed in such proximity-effect tunnel junctions.

Despite the failure of earlier efforts, the new attempt¹ by workers at Philips Research Laboratories in the Netherlands seems to be relatively successful. These researchers covered a 20-nanometre layer of silver previously deposited on a thin film of $YBa_2Cu_3O_7$ superconductor with a further layer of insulating aluminium oxide followed by a top electrode made of lead (Fig. 1). The electrical characteristics of the junction (Fig. 2) are notable for the very low level of 'leakage' current, the current that persists at low voltages. They are also close to what one expects from conventional 'BCS' theory of

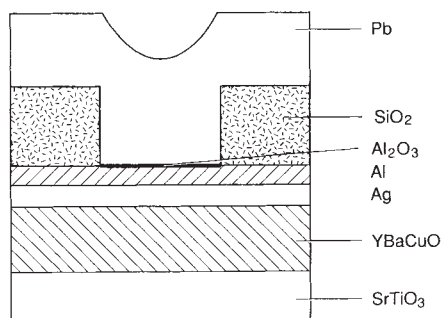


FIG. 1 Proximity-effect tunnel junction fabricated by Gijs *et al.*

RÉSUMÉ

Colour coding

THE theory that parasitic infection underlies the evolution of brightly coloured plumage in male birds takes a fresh blow with a report by S. G. Johnson in the latest issue of *Evolutionary Ecology* (**5**, 52–62; 1991). W. D. Hamilton and M. Zuk's 'revealing handicap model', proposed in 1982, holds that females choose to mate with brightly coloured males because showiness correlates with resistance to parasites; the model predicts that males from species subject to more intense parasitism will be more extravagantly plumaged. The results of early studies with North American and European passerines conformed to this prediction, but reanalysis by A. F. Read and P. H. Harvey, using brightness scores supplied by ornithologists unaware of the theory, found no relationship. Johnson now confirms Read and Harvey's result for a larger sample of North American species, and shows that variation in male brightness is more strongly associated with phylogeny and the risk of predation.

Cool for clouds

THE discovery last year of cold clouds of hydrogen in the distant Universe caused quite a stir: distant clouds are well known, but they are hot and are thought to be the size of galaxies, fitting in with popular notions about the evolving Universe. But for these clouds to be so cold, despite being subjected to intense ionizing radiation from even more distant quasars, they would have to be a million times smaller, so making an entirely new component in our construction of the distant Universe (see J. Peacock's News and Views article, *Nature* **349**, 190–191; 1991). Not necessarily, claim R.C. Duncan, E.T. Vishniac and J.P. Ostriker (*Astrophys. J.* **368**, L1–L5; 1991): cold means small only if the clouds are at equilibrium. These authors and others have previously argued that distant clouds should be expanding (or inflating) and that expansion efficiently cools them despite the ionizing radiation. This allows the cold clouds to be as large as their hot companions, thereby rescuing the old picture.

Out of sequence

EXONS can be spliced together in the 'wrong' order. This startling finding, reported by J. M. Nigro *et al.* (*Cell* **64**, 607–613; 1991), stems from experiments to identify the order of exons within a potential tumour-suppressor gene in humans. The mechanism of the aberrant splicing reactions is unclear, but could involve *trans*-splicing (which has not yet been observed in mammalian systems *in vivo*) or the looping back of a downstream exon so that its 3' boundary fortuitously becomes aligned with the 5' boundary of the upstream exon. Whatever the mechanism, the conclusion must be that pre-messenger RNA processing is less precise than previously thought, which in turn bears upon the way genetic diversity was introduced into an 'RNA world'.

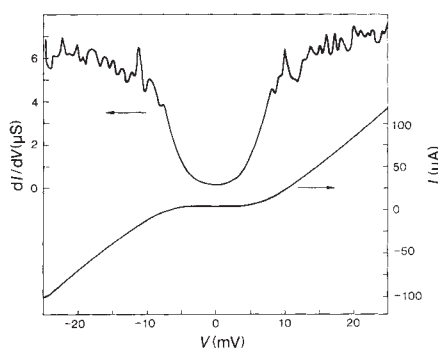


FIG. 2 Current-voltage (I/V) and differential (dI/dV) characteristics of the junction prepared by Gijs *et al.*

superconductivity. On the other hand, the energy gap in the metal overlayer reported by the Dutch workers (which reflects the strength of binding between the electrons in the superconducting electron pairs) is about half that reported by those using cruder tunnel junctions. The superconducting energy gap disappeared when the temperature was raised to 20 K, perhaps owing to changes in the proximity effect.

Earlier work⁴ had shown that the proximity coupling is very anisotropic, as one might expect in these quasi-two-dimensional layered materials. The coupling is strong when the gold or silver layer is deposited on the edges of the famous CuO_2 layers in these high temperature superconductors and essentially nonexistent when the deposition is on the faces. The proximity coupling is also much weaker than the ideal limit permitted by theory. The films of the Dutch group had grains with both types of faces exposed at the surface, so there is no contradiction.

Theorists at Moscow State University recently argued⁵ that reduced proximity coupling is to be expected at the interface of the oxide superconductors and conventional normal metals such as gold or silver. The difference in electron density between the high-temperature superconductors and such normal metals leads to reflection of the electrons at the interface (and hence a boundary resistance) that inherently reduces the proximity coupling, and concomitantly the induced energy gap as well. The physics here is well understood, but its implications had not been fully appreciated.

Surface scientists are not without their view in these matters: they are of a contrary frame of mind. Wells *et al.*² sought evidence for an induced superconducting energy gap in gold proximity layers deposited on top of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ single crystals and on $\text{YBa}_2\text{Cu}_3\text{O}_7$ thin films. In the case of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$, the coupling was to the flat surface of the CuO_2 planes, and in the case of $\text{YBa}_2\text{Cu}_3\text{O}_7$ it was to the edges. No evidence for a gap was seen in either case. The superconducting energy gap has been clearly observed by several groups in high-resolution photoemission spectroscopic studies of clean cleaved surfaces of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$.

Even if good tunnel junctions evolve from

this work, technologists will still have to wait for commercial junctions. The experiments described here involves junctions with low-temperature superconducting counterelectrodes — usually made of lead, which has a transition temperature of 7.2 K. A junction with both electrodes made from a high-temperature superconductor has yet to be reported, although many groups are feverishly working on the problem.

And there may be some competition in the wings. Workers at AT&T Bell Laboratories have recently reported³ a tunnel junction with remarkably good characteristics formed entirely from $(\text{Ba,K})\text{BiO}_3$. These bismuth-oxide superconductors are moderately low-temperature cousins to the more famous copper-oxides. The $(\text{Ba,K})\text{BiO}_3$ system has been shown to have transition tem-

peratures as high as 30 K. It is isotropic, and not quasi-two-dimensional as are the copper oxides. However, in this new work, the junction was actually made with a point contact formed by pressing a $(\text{Ba,K})\text{BiO}_3$ bulk crystal onto a thin film of the same material. Thus there is also more work to do here before the technologists can give them a try. □

M. R. Beasley is in the Department of Applied Physics, Stanford University, Stanford, California 94305, USA.

1. Gijs, M.A.M., Scholten, D., van Rooy, Th. & Gerrits, A.M. *Appl. Phys. Lett.* **67**, 2600–2602 (1990).
2. Wells, B. O. *et al. Phys. Rev. Lett.* **65**, 3056–3059 (1991).
3. Targellis, A.N. *et al. Appl. Phys. Lett.* **58**, 95 (1991).
4. Akoh, H. *et al. Appl. Phys. Lett.* **57**, 1152–1154 (1991).
5. Kuprianov, M. Yu & Likharev, K. K. *Proc. 1990 Appl. Supercond. Conf.* (in the press).

ENZYMOLOGY

Radical copper in oxidases

A. J. Thomson

OXIDASES containing a mononuclear copper site catalyse the two-electron oxidation of substrate by di-oxygen, which is thereby reduced to hydrogen peroxide. For example, galactose oxidase, an extracellular enzyme from the fungus *Dactylium dendroides*, oxidizes primary alcohols¹, whereas amine oxidases such as methylamine oxidase found in the Gram-positive bacterium *Arthrobacter P1* (ref. 2) and in bovine serum, oxidize primary amines. Both enzymes yield aldehydes. The mechanism of these enzymes poses the question as to how a single copper centre capable of undergoing only a one-electron change $\text{Cu(I)}/\text{Cu(II)}$ catalyses a two-electron transfer from organic substrate to O_2 .

In both cases the answer seems to be by the use of a redox active organic cofactor in close proximity to the copper centre. The crystal structure of galactose oxidase at 1.7 Å resolution, reported on page 87 of this issue³, shows that the coordination site of the copper ion contains two histidine and two tyrosinate residues. One of the latter, Tyr 272, forms a thioether bond to Cys 228 and also has a π -stacking interaction with the indole ring of nearby Try 290. Spectroscopic evidence⁴ indicates that the active state of the enzyme can store two oxidizing equivalents, one Cu(II) ion and the other as a radical cation (now, as a result of the X-ray data, identified as that of the covalently modified Tyr 272 side chain). The copper-containing methylamine oxidase also apparently contains a redox active organic cofactor that has been characterized by degradation studies as 6-hydroxydopa quinone⁶. Studies using electron para-magnetic resonance reported in January⁷ reveal the presence of a Cu(I) -semiquinone which may be the catalytic intermediate that undergoes a two-electron oxidation by di-oxygen.

It had long been suspected that both galac-

tose oxidase and methylamine oxidase contain a redox active organic cofactor in addition to copper ion, but its chemical nature remained controversial. Reports from the laboratory of J.A. Duine claimed the presence of covalently bound PQQ (pyrroloquinoline quinone) in both enzymes^{8–10} (Fig. 1). This molecule, non-covalently bound, is the only cofactor required by the enzyme methanol dehydrogenase found in methylotrophic bacteria. In this case the absolute requirement for this cofactor can be demonstrated by the recovery of enzyme activity on addition of PQQ to apo-protein.

Unambiguous identification of the presence of a covalently bound form of PQQ has however been difficult to obtain. A remarkably wide range of enzymes has been reported to contain PQQ, including lipoxigenase¹⁰, lysyl oxidase¹¹ and dopamine β -

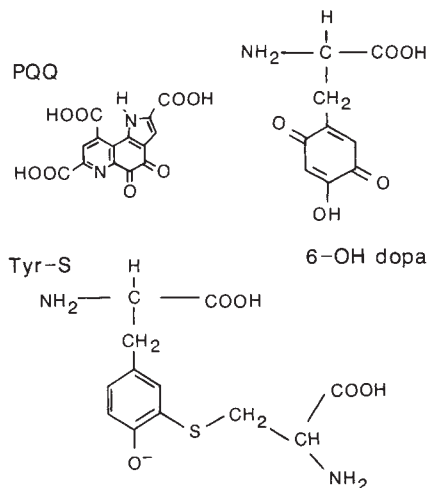


FIG. 1 The chemical structures of the redox active organic cofactors found in methanol dehydrogenase (PQQ), in methylamine oxidase (6-OH dopa) and in galactose oxidase (Tyr-S).

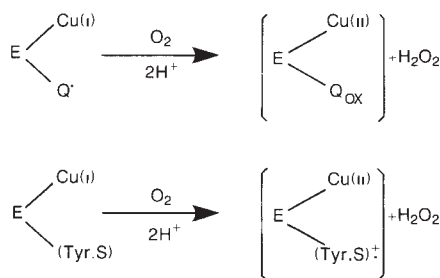


FIG. 2 The nature of the catalytic intermediate generated by di-oxygen in (top) methylamine oxidase, where $Q \equiv$ 6-hydroxydopa quinone; and (bottom) galactose oxidase, where Tyr-S represents Tyr covalently linked to the thiol side-chain of cysteine.

hydroxylase¹². But these claims have relied primarily on the colorimetric identification of the formation of a phenylhydrazone derivative by reaction of the carbonyl function of *ortho* diquinone with phenylhydrazine. But this reaction is not specific for PQQ; the covalent binding of the cofactor to the protein requires hydrolysis in order to release a low-molecular-weight fragment for analysis. Recently, mild enzymatic proteolysis of methylamine oxidase, followed by mass spectrometric and nuclear magnetic resonance studies, has identified the cofactor of methylamine oxidase as 6-hydroxydopa and not PQQ⁶. It remains to be seen whether lipoxxygenase, lysyl oxidase and dopamine β -hydroxylase also contain 6-hydroxydopa.

The Cu(I)-semiquinone state of amine oxidase has not been detected previously because this state is apparently in thermal equilibrium with Cu(II)-reduced quinone. Internal electron transfer from copper to semiquinone is favoured by low temperatures, and since previous work used low-temperature electron paramagnetic resonance spectroscopy to study the copper centre, the semiquinone form, was missed. The new data on the structure of galactose oxidase provides yet another example of a metalloprotein in which a tyrosine radical appears to play a catalytic role. This list contains ribonucleotide reductase¹³ and prostaglandin H synthase¹⁴.

A comparison can now be made between the structures of the catalytic intermediates

of galactose oxidase and methylamine oxidase generated by oxygen (Fig. 2). In the case of galactose oxidase the reactive intermediate contains Cu(II) and the modified tyrosine radical, whereas in methylamine oxidase it is Cu(II) and the quinone form of 6-hydroxydopa. The radical is generated in the reduced state in one case and the oxidized state in the other. It is clear from the X-ray structure that, in the case of galactose oxidase, the

tyrosine radical is liganded to the copper ion and hence may function as a concerted two-electron oxidant. It is not yet known whether the quinone in methylamine oxidase is bonded to copper nor whether substrates interact at the metal site. Answers to these questions should soon be forthcoming. $\square$

A. J. Thomson is in the School of Chemical Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

EVOLUTIONARY ECOLOGY

How to live like a mammal

Paul H. Harvey and Sean Nee

"LIVE fast, die young and leave a good looking corpse" is a life-history strategy seemingly followed by some mammals such as mice. Others, like ourselves, prolong each life-history stage. Recently, considerable

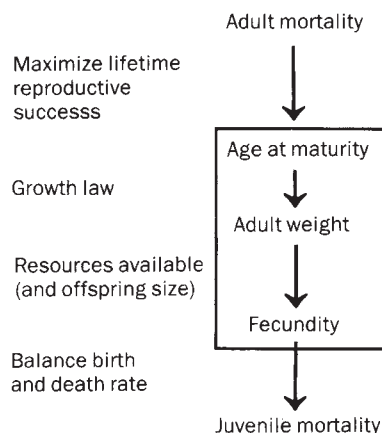


FIG. 1 Charnov's view of the causal chain in mammalian life-history evolution. Adult mortality is determined by the environment. Mammals mature at the age which maximizes lifetime reproductive success. A growth law determines adult size as a function of age at maturity. At maturity, 'the energy which would have gone into growth is channelled into the production of offspring. In the long term, the birth rate must equal the death rate in natural populations. That equality is ensured by juvenile mortality being density dependent: at low population densities fecundity exceeds mortality, but at high density mortality exceeds fecundity.

effort has been invested into describing and explaining various features of mammalian life-history diversity. Now, in a bold initiative, Charnov¹ attempts to integrate the evidence and produce the first comprehensive model of life-history evolution for female mammals (Fig. 1).

Unravelling the evolution of life histories is particularly difficult because so many variables are involved, and so many more impinge upon the system. For example, fecundity has several components including time to maturity, litter size and the interval between litters. Mortality rates change with age: neonatal mortality is usually much higher than juvenile or early-adult mortality, while older mammals senesce. Add to this

the fact that some components of fecundity and mortality are dependent on population density, and things look very complicated indeed.

Nevertheless, hope for analysing the variation in terms of only a few underlying parameters comes from the finding that the components of life histories tend to covary; for example, larger mammals have longer gestation lengths, later ages at maturity, longer intervals between litters, smaller litters, and lower mortality rates as both infants and adults². Such covariation implies that all life histories may be determined by some key variable. Many possibilities have been suggested, including brain size^{3,4}, metabolic rate⁵, and even an elusive 'periodonger' which entrains the timing of life-history events to body weight⁶. In fact, on close examination it seems that body weight is at least as good a predictor of life-history variation as any alternative yet proposed⁷.

Size and energy seem inextricably linked. The rate at which mammals put on weight after weaning is proportional to their body weight raised to the 0.75 power⁸. If, when mammals stop growing, they divert the energy available to them from growth into reproduction, then larger-bodied animals will have more to invest in their offspring per unit time. But size and time are also inextricably linked. It takes longer to grow to a large size than a small one. So there is an obvious disadvantage for an animal in continuing to grow until it is possible to invest vast amounts of energy in reproduction: she might die first! This trade-off is the core of Charnov's optimality model.

Charnov suggests that 'extrinsic' mortality rates around that time of maturity determine when mammals stop growing, and thereby dictate adult body size. Extrinsic forces of mortality are misfortunes such as lightning bolts, predation and falling missiles. The new theory is unconventional in that it ignores 'intrinsic' forces of mortality, which part be determined by reproductive decisions. For example, if energy was diverted from maintenance to reproduction, the mortality rate might be increased. Charnov suggests audaciously that such intrinsic forces of mortality may be irrelevant for understand-

1. Amaral, D., Bernstein, L., Morse, D. & Horecker, B.L. *J. biol. Chem.* **238**, 2281–2284 (1963).
2. Mondovi, B. (ed.) *Structure and Function of Amine Oxidases* (CRC Press, Boca Raton, 1985).
3. Ito, N. *et al. Nature* **350**, 87–90 (1991).
4. Whittaker, M.M., DeVito, V.L., Asher, S.A. & Whittaker, J.W. *J. biol. Chem.* **264**, 7104 (1989).
5. Whittaker, M.M. & Whittaker, J.W. *J. biol. Chem.* **263**, 6074–6080 (1988).
6. James, S.M. *et al. Science* **248**, 981–987 (1990).
7. Dooley, D.M. *et al. Nature* **349**, 262–264 (1991).
8. Ameyana, M., Hayashi, M., Matsushita, K., Shinagawa, E. & Adachi, O. *Agric. biol. Chem.* **48**, 561–570 (1984).
9. Lobenstein-Verbeek, C.L., Jongejan, J.A., Frank, J. & Duine, J.A. *FEBS Lett.* **170**, 305–309 (1984).
10. Van der Meer, R.A. & Duine, J.A. *FEBS Lett.* **235**, 194–200 (1988).
11. Williamson, P.R., Moog, R.S., Dooley, D.M. & Kagan, H.M. *J. biol. Chem.* **261**, 16302–16305 (1986).
12. Van der Meer, R.A., Jongejan, J.A. & Duine, J.A. *FEBS Lett.* **231**, 303–307 (1988).
13. Norabund, P., Sjöberg, B.-M. & Eklund, H. *Nature* **345**, 593–598 (1990).
14. Kartheim, R., Dietz, R., Nastainczyk, B. & Ruf, H.H. *Eur. J. Biochem.* **171**, 313–320 (1988).

ing the comparative patterns of life-history diversity.

Such a framework leaves two major components of the system to be considered: neonatal and juvenile mortality, and the partitioning of available energy into more and smaller or fewer and larger offspring. In accord with empirical evidence^{9,10}, Charnov considers juvenile mortality to be determined by density-dependent factors so as to keep the birth and death rate of the population balanced.

Larger-bodied mammals not only have more energy to put into reproduction, they also wean heavier young and so their fecundity in terms of young produced per unit time is lower than for small-bodied species. Why should larger-bodied species wean heavier young? An interesting feature of Charnov's model is that optimal adult body size is independent of weaning size, so offspring that are weaned heavier will spend less time growing as juveniles and, as a consequence, are more likely to survive to breed themselves.

From these assumptions, Charnov shows that the time between weaning and maturity should increase with the +0.25 power of body size, while litter size per unit time, and both adult and juvenile mortality rates, should scale with the -0.25 power of body size, in accord with the evidence. The theory also predicts that the product of age at maturity and the instantaneous adult mortality rate should be about 0.7. Data collected from 26 mammal species¹¹ fit this expectation.

Work on the covariation of life-history traits among mammal species suggests that size is not everything: life-history variables often covary when size effects are held constant by partial correlation⁷. For example, those species with an early age at maturity for their body size are also the ones with high adult mortality rates for their body size. Charnov's model can account for such pat-

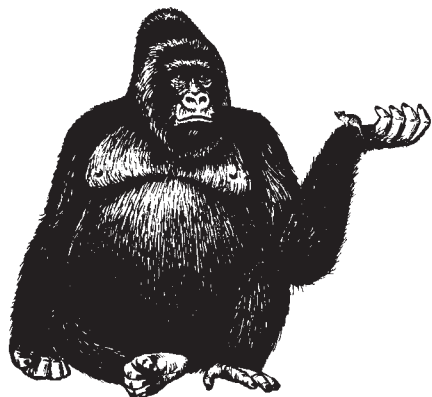


FIG. 2 The smallest and largest primate species, a gorilla *Gorilla gorilla* and a mouse lemur *Microcebus murinus*, have very different life-history characteristics.

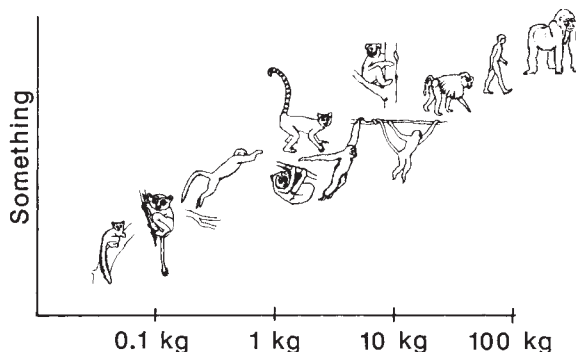


FIG. 3 Life-history variation is well predicted by body weight across primates; 'Something' in the figure refers to the timing of crucial events in the life history, such as gestation, weaning, juvenile period or lifespan. Relationships are typically linear when both the size and timing axes are logarithmically scaled. (Reproduced from ref. 12.)

terns when it drops the assumption that all mammals grow along the same trajectory: even though growth rate increases with body size raised to the +0.75 power within species, species differ in their growth rates at a given body weight. Those with higher mass specific growth rates would, under Charnov's model, have higher adult mortality rates and earlier ages at first reproduction for a given adult body weight. Other parameters which may vary across species, including the efficiency of transfer of resources from potential growth to the production of offspring, also lead to biologically observed and Charnov-predicted correlations between life-history variables independently of size.

Charnov's framework for the study of life-history variation provides an integrated set of testable predictions such as have long been needed by those trying to make sense of mammalian life-history diversity. The model itself will, no doubt, evolve as its component parts are tested. It will be vital, however, to examine how changes in one part ramify through the whole framework: changing one assumption in such an integrated model will have many consequences. □

Paul H. Harvey and Sean Nee are in the AFRC Unit of Ecology and Behaviour, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Charnov, E. L. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1134–1137 (1991).
2. Western, D. Afr. J. Ecol. **17**, 185–204 (1979).
3. Sacher, G. A. in *CIBA Foundation Symposium on the Lifespan of Animals* (eds Wolstenholme, G. E. W. & O'Connor, M.) 115–133 (Little, Brown and Co., Boston, 1959).
4. Sacher, G. A. & Staffeldt, E. F. *Am. Nat.* **108**, 593–616 (1974).
5. McNab, B. K. *Ecol. Monogr.* **56**, 1–19 (1966).
6. Lindstedt, S. L. & Swain, S. D. in *Evolution of Life Histories of Mammals: Theory and Pattern* (ed. Boyce, M. S.) 93–106 (Yale Univ. Press, New Haven, 1988).
7. Harvey, P. H., Read, A. F. & Promislow, D. E. L. *Oxford Surveys in Evolutionary Biology* **6**, 13–31 (1989).
8. Lavigne, D. M. J. *Anim. Ecol.* **51**, 195–206 (1982).
9. Fowler, D. C. *Ecology* **62**, 602–610 (1981).
10. Sinclair, A. R. E. in *Ecological Concepts* (ed. Cherritt, J. M.) 197–241 (Blackwell Scientific, Oxford, 1969).
11. Millar, J. S. & Zammuto, R. M. *Ecology* **64**, 631–635 (1983).
12. Fleagle, J. G. in *Size and Scaling in Primate Biology* (ed. Jungers, W. L.) 1–20 (Plenum, 1985).

Safe houses

THE hovercraft is an elegant device marred by a serious flaw. Its resilient cushion of compressed air would be an ideal frictionless support, but for the copious and continuous air-flow needed to replenish the steady leakage around the periphery.

So Daedalus is inventing a liquid seal for hovercraft. Modern pumps and compressors are often lubricated with magnetic oil. It contains fine iron or iron-oxide particles which hold it securely in a suitably magnetized bearing even against considerable gas pressure. Daedalus's magnetically sealed hovercraft has a peripheral skirt of flexible, ferrite-loaded magnetized rubber, to which a thick magnetic oil permanently adheres. On a smooth surface, it merely needs to be pumped up until its weight is carried by the air-cushion within. It can then glide frictionlessly along with no further pumping. The oil-seal around the edge cannot be breached by internal pressure, nor wiped away by movement.

At first Daedalus saw this novel craft as an ideal heavy lorry, with no destructive rubber tyres to damage the roads. But he soon realized that it would still have to be propelled, steered and braked by noisy and cumbersome aero-engines. He is now developing it as a neat and simple support for heavy objects that occasionally must be moved around: cranes, freight containers, portable cabins, even grand pianos and filing cabinets. Fitted with the DREADCO integral hover base, they could be pumped up to neutral buoyancy, manoeuvred effortlessly around, then deflated again to rest firmly in their new location. A vast amount of undignified lifting and shoving, or manipulation with low-loaders and forklift trucks, could be avoided.

Even architecture could benefit. A modern building could be levitated above its concrete-raft foundation with a fraction of an atmosphere of retained air-pressure. Office developers would welcome the chance to shift their monstrosities around, always seeking the currently most fashionable and highest rent locations. But the real benefit would be earthquake protection. For this purpose the magnetized-rubber skirt, as well as being able to slide sideways on its base, would be made deep enough to permit a degree of vertical travel. At the first warning or small preliminary tremor, automatic compressors could pump up the 'hover building' to resilient and frictionless suspension on this private air-cushion. It would then be totally decoupled from lateral or vertical ground-movements which would be fatal to rigidly anchored buildings. When the danger was safely past, the air-cushion would be deflated again, leaving the building erect and stable on its original site, smugly intact amid the ruins of less ingenious structures.

David Jones

Pollution and the Persian Gulf

SIR — The waters of the Persian Gulf are not well confined, despite the narrowness of the Strait of Hormuz, which links the Gulf to the Arabian Sea. Renewal of Gulf waters through the strait is rapid, owing to missing sills and strong thermohaline circulation. The average annual evaporation rate of the Persian Gulf is about 1.5 m yr^{-1} (ref. 1). Estimates of relevant time scales for overturning, residence and renewal of Gulf waters range from less than one year to about four years¹⁻³. Outflowing Gulf waters (about $3,000 \text{ km}^3 \text{ yr}^{-1}$) pass through the Strait of Hormuz as a bottom current along the depths of the Arabian coast⁴. The compensating inflow ($0.1-0.2 \text{ m s}^{-1}$) occurs at the surface and mainly along the Iranian side². The outflow from the Gulf then follows the bottom slope of the Gulf of Oman and spreads through the Arabian Sea at depths of about 200–600 m (ref. 5). So any environmental impact on the Gulf waters is rapidly exported to the Arabian Sea and then to the Indian Ocean.

The Gulf War is causing increased pollution of Gulf waters and its impacts cannot be assessed directly. I suggest monitoring the outflow of the Gulf waters in the Strait of

Hormuz off Musandam and in the western Gulf of Oman. If immediate action is taken, the first measurements will record pre-war conditions. Subsequent shifts in levels of contamination of the outflowing waters might be used to assess the environmental impact of the war on both the Gulf waters and those of adjacent seas.

Funding for this monitoring programme should come from international agencies; management of such monitoring should be undertaken by a nongovernmental organization or a national institution in a country not involved in the war.

MARTIN BOHLE-CARBONELL

*Institut für Meereskunde,
Tropelwitzstr. 7,
D-2000 Hamburg 54,
Germany*

1. Koske, P. H. 'Meteor' Forschungsergebnisse All (1972).
2. Hunter J. R., *Unesco Reports in Marine Science* **28**, (1984).
3. Hassan, E. M. & Hassan, H. M. *Arab Gulf J. Sci. Res.* **7**, 93–109 (1989).
4. Ziegenbein, J. 'Meteor' Forschungsergebnisse Al (1966).
5. Düing, W. & Schwill, W. D. 'Meteor' Forschungsergebnisse A3 (1967).

Radiation doses and cancer

SIR — In February 1990, the Soviet government allowed the release of data on the radiation doses and cancer rates to the workers in the first Soviet atom-bomb facility, near Chelyabinsk. Nikipelov *et al.* published in *Priroda (Nature)*¹ the radiation doses for each year, averaged over the workforce, and the cancer rates for high- and low-dose groups. Unfortunately, they did not report the number in the workforce.

Pending release of the full data, we have deduced an estimate of the total number of workers, and hence the workforce, from the errors σ assigned to the cancer rates π reported by Nikipelov *et al.* We assume that the authors used standard statistical formulae, and obtain $n = \pi \cdot (1-\pi)/\sigma^2$. An entry in their Table 3, for example, which we reproduce here, gives a mortality from cancer of $5.7\% \pm 0.6\%$ for those workers at the reactor, employed before 1958, who received less than 100 rem. From $\pi=0.057$ and $\sigma=0.006$, we estimate that there must have been at least 1,500 people in the workforce

at the reactor. By combining several estimates, we determine that the number of people employed before 1958 was 2,000 at the reactor and 4,600 at the plutonium separation plant. If we allow for an average work period of five out of the 10 years, we should halve these numbers to obtain the workforce at any one time. This work period of five years may be an underestimate, even for the early years, as highly exposed workers were probably reassigned to jobs with lower exposure, rather than being dismissed, and hence remained in the listed work-force.

The collective radiation dose, then, is simply the workforce multiplied by the sum of the yearly averaged radiation doses. From Table 1 of ref. 1, we find 260,000 person rem for the reactor, and 1.2 million person rem for the processing plant. To estimate the number of cancers that were caused by these doses, we take from the table the 3.6% difference of the cancer rates for two dose groups: greater and less than 100 rem, multiplied by the 3,100 people that we estimate received more than 100 rem. We assume that the average age of the workers was 25 years. These people would now be 65 years old, and some of their cancers would not yet

have been expressed. We therefore multiply the number of cases by a factor of 3.1 to obtain the number of cases by the end of life. (Estimate from the US statistical abstract, 1989.) This factor is sensitive to the average age of the workers; from the same data we find that the lifetime cancer mortality in the Soviet Union is a reasonable 15%; less than 25% in the United States, where life expectancy is greater. We find 67 ± 24 lifetime cases from reactor exposures, and 242 ± 46 from the reprocessing facility for a total of 310 ± 52 . This figure should be increased further by adding the smaller number of cancers among the equal number of people who accumulated less than 100 rem. Assuming that these received an average of 30 rem, and that linearity applies, another $1/4$ or 80 cases should probably be attributed to the lower exposures for a total of 400. Dividing by the collective dose, we find >3 'excess deaths' per 10,000 person rem.

This risk of cancer is smaller than the risk for atom-bomb survivors with similar doses (for example, the BEIR-V report² estimates about 8 deaths per 10,000 person rem). But animal data show that spreading the doses over time can reduce the risk by a factor between 2 and 10. The atom-bomb doses were almost instantaneous, whereas the doses at Chelyabinsk, although peaking when the workers installed radioactive equipment, were spread out in time. This comparison of the effects from Chelyabinsk and the atom-bombs confirms that there is some reduction factor. Stewart and Kneale have recently suggested³ that using the Hiroshima–Nagasaki cohort understates the risk because this cohort is one of healthy survivors. This argument would not apply to the Chelyabinsk cohort. Finally, the total collective dose at Chelyabinsk exceeds that for the life span study (LSS) cohort from Hiroshima and Nagasaki, and the number of people in the former cohort is less than 10,000 compared with 76,000 in the latter¹. Therefore, the excess deaths from cancer will be far easier to distinguish from background.

Because the whole world is interested in these results, we call for international support to help scientists in the Soviet Union understand these issues, both with experts and with money.

ALEXANDER SHLYAKHTER
RICHARD WILSON

*Department of Physics,
Harvard University,
Cambridge,
Massachusetts 02138, USA*

LEVEL OF CANCER MORTALITY AMONG THE EMPLOYEES OF THE CHELYABINSK FACILITY DURING 40 YEARS OF OBSERVATION

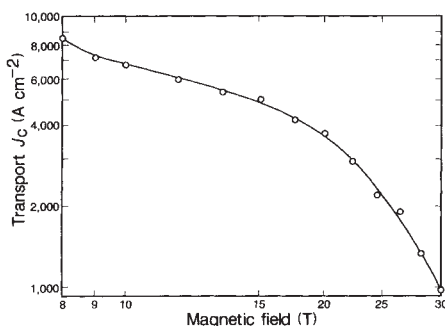
Plant	Total gamma dose, rem		Maximum year dose, rem	
	<100	>100	<25	>25
Reactor	5.7 ± 0.6	9.4 ± 1.2	5.9 ± 0.7	8.7 ± 1.1
Plutonium separation	4.3 ± 0.4	8.1 ± 0.6	4.2 ± 0.5	7.7 ± 0.5
Total	4.8 ± 0.4	8.4 ± 0.5	4.9 ± 0.4	7.9 ± 0.5

Figures are percentages of the number of people starting work before 1958 (from ref: 1).

1. Nikipelov, B.V., Lizlov, A.F. & Koshurnikova, N.A. *Priroda* pp 30–38 (February 1990).
2. National Research Council Committee on the Biological Effects of Ionizing Radiation (BEIR V) *Health Effects of Exposure to Low Levels of Ionizing Radiation* (National Academy, Washington DC, 1990).
3. Stewart A.M. & Kneale, G. *Health Phys.* **58**, 729–735 (1990).
4. Shimizu, Y., Kato, H. & Schull, W.J. *Life Span Study Report No. 11, Part 2*. (Radiation Effects Research Foundation, 1988).

Current record in superconductors

SIR — The achievement of high transport critical currents in bulk high- T_c superconductors at high magnetic fields and temperatures is necessary for many applications of these new materials¹. The recently developed melt-growth process^{2–4} offers the potential for obtaining greatly enhanced transport critical current density, J_c , compared with sintered high- T_c materials¹. We show here that high transport J_c can be achieved in bulk melt-grown $\text{YBa}_2\text{Cu}_3\text{O}_7$ (ref. 4) at magnetic fields up to 30 T, at liquid-nitrogen temperature. (This is well above the irreversibility field, typically



Transport critical current density versus magnetic field for bulk grain-oriented $\text{YBa}_2\text{Cu}_3\text{O}_7$ at liquid-nitrogen temperature, with $\mathbf{B}$ perpendicular to the c -axis.

quoted as ~6 T for $\text{YBa}_2\text{Cu}_3\text{O}_7$ at 77 K for field along the c -axis.) As shown in the figure, the transport J_c for mutually perpendicular current and magnetic field in the a - b plane was $1\text{--}8\text{ kA cm}^{-2}$ up to 30 T. We think that these are the highest d.c. transport J_c values yet reported for bulk $\text{YBa}_2\text{Cu}_3\text{O}_7$ at 77 K and high magnetic fields. Such a high J_c at these high fields has not been obtained in bulk bismuth- and thallium-based high T_c materials at 77 K because of the strong thermally activated flux creep in these systems.

At a lower temperature of 4.2 K, the critical current exceeded the current capacity (200 A) of our vapour-cooled current leads, and we are able to determine only a lower bound for the transport J_c of bulk $\text{YBa}_2\text{Cu}_3\text{O}_7$ at 4.2 K of $>22\text{ kA cm}^{-2}$ at 30 T. These data suggest, however, that in the intermediate temperature range between 20 and 40 K the transport J_c at high fields over 30 T may well reach practical levels ($>10^4\text{ A cm}^{-2}$). These results bode well for applications such as current leads operating between liquid nitrogen and liquid helium temperature. They also provide motivation for the difficult task of developing long conductors made of high- T_c material for high-temperature magnet applications.

The low-contact-heating measurements necessary to measure such high J_c s were made with high-quality silver contacts⁵ and a high-current sample mount. An electric field criterion of $10\text{ }\mu\text{V cm}^{-1}$ was used to determine the critical current⁶. On cycling the

field between 8 T and 30 T, the critical current was reversible to within the experimental precision of $\pm 5\%$.

The sample dimensions for these transport J_c measurements were $\sim 6 \times 1.7 \times 0.6\text{ mm}$, with $\sim 3\text{ mm}$ between the voltage taps. The samples were fabricated using a liquid-phase processing method described in detail elsewhere⁷. In this process, sintered bars of $\text{YBa}_2\text{Cu}_3\text{O}_7$ are melted at $1,100\text{ }^\circ\text{C}$ for 10–15 min to decompose the compound into Y_2BaCuO_5 and liquid. The melt is then cooled slowly ($1\text{--}2\text{ }^\circ\text{C h}^{-1}$) from $1,025$ to $925\text{ }^\circ\text{C}$, through the peritectic temperature³. Following the liquid-phase process, the samples were annealed in oxygen for 24 h at each of $500\text{ }^\circ\text{C}$ and $400\text{ }^\circ\text{C}$. Further details on the measurements will be published elsewhere.

J. W. Ekin

Electromagnetic Technology Division,
National Institute of Standards and
Technology,
Boulder,
Colorado 80303, USA

K. SALAMA

V. SELVAMANICKAM

Texas Center for Superconductivity and
Mechanical Engineering Department,
University of Houston,
Houston,
Texas 77004, USA

1. Ekin, J. W. *Adv. Ceram. Mater.* **2**, 586–592 (1987).
2. Jin, S., et al. *Appl. Phys. Lett.* **52**, 2074–2076 (1988).
3. Murakami, M., Morita, D., K. & Miyamoto, M. *Jap. J. Appl. Phys.* **28**, 1189–1194 (1989).
4. Salama, K., Selvamannickam, V., Gao, L. & Sun, K. *Appl. Phys. Lett.* **54**, 2352–2354 (1989).
5. Ekin, J. W. et al. *Appl. Phys. Lett.* **52**, 1819–1821 (1988).
6. Ekin, J. W. *Appl. Phys. Lett.* **55**, 905–907 (1989).
7. Selvamannickam, V. & Salama, K., *Proc. MRS Meet.* **169**, 279 (1990).

Gas hazard on Vulcano Island

SIR — Baubron *et al.*¹ used data acquired during a single survey of CO_2 from soils around Vulcano Porto to define the gas hazard and the origin of the examined gases. Other groups, coordinated by the National Group for Volcanology, the Institute of Geochemistry of Fluids of the CNR and the Institute of Mineralogy Petrography and Geochemistry of the University of Palermo, have been carrying out geochemical and gas-hazard monitoring since 1977 on the island of Vulcano^{2–8}.

Between 1984 and 1990, 45 soil CO_2 surveys were carried out in the area between the base of the active cone of the Fossa and the walls of the surrounding caldera, where the gas hazard is higher because of terrain morphology and population density. In these surveys, CO_2 fluxes from the ground were measured according to the method developed in ref. 3. Soil CO_2 measurements were carried out on a sampling grid of 50 points distributed over an area of 2.2 km^2 . Unlike Baubron *et al.*, a homogeneous grid was used because degassing from soils is

spatially dishomogeneous, being mainly controlled by tectonics⁹.

A wide range of fluxes, up to three orders of magnitude, has been measured. Maximum output was normally found at the base of the Fossa cone, close to the Faraglioni area. Any observed increase in CO_2 output is associated with an enlargement of the higher exhaling area. Temporal variations in unitary flux show different peaks at different amplitudes (Fig. 1). Two separate moments of maximum degassing in May 1985 and October 1987, can be identified. Significant variations in the activity (increased maximum temperature, seismicity and compositional variations, in fumaroles) were observed in both periods^{5,7}.

From June to September 1988, the frequency of observation and sampling was increased as a response to the increase in activity. In particular, measurements of soil CO_2 flux were intensified to four per month to check the gas hazard during the summer, when more than 15,000 tourists visit the island.

The CO_2 flux showed an increase from the first 10 days of August, up to a maximum value in October. Increased CO_2 output from the ground was again synchronous with variations observed in the composition and output of fumarole gases in the crater area (increased He, H, CO, reducing capacity, steam output)^{4,5}. Increased seismic activity was also recorded in the last 10 days of August⁴.

Our data clearly reveal that, although some peaks were recorded in the summer and autumn of 1988, the CO_2 output shows a general decreasing trend from October 1987 to March 1990. Furthermore, the mean values of eight continuous measuring points installed since the end of 1988 are relatively low, confirming that the gas hazard during 1988 was reduced. Thus, any reliable evaluation of gas hazard must be based on systematic monitoring, as isolated or sporadic recordings can lead to erroneous conclusions.

Our average unitary flux data allowed us to estimate CO_2 output from the area examined as 200 tons per day in September–October 1988, a value comparable to that emitted from the crater fumaroles⁴. In the same period, Baubron *et al.*¹ evaluated soil CO_2 output at 30 tons per day. As the size of the areas to which they refer is not indicated,

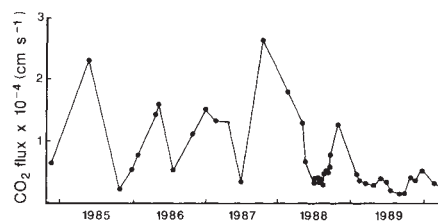


FIG. 1 Unitary flux of CO_2 from soil recorded at Vulcano Island from 1984 to March 1990. Despite an increase in activity observed since October 1987, CO_2 flux tends to decrease.

their data cannot be immediately compared with ours. However, they report concentrations between 6.5 and 87% vol at a depth of 80 cm and a flux ranging between 0.23 and 2.64 $1 \text{ m}^{-2} \text{ h}^{-1}$. Assuming that CO_2 flux is exclusively supported by diffusion, that the concentration gradient is constant, that the diffusion coefficient is 0.12 $\text{cm}^2 \text{ s}^{-1}$, and that soil porosity is 0.33 (ref. 3), specific CO_2 flux values between 1.64 and 20.2 $1 \text{ m}^{-2} \text{ h}^{-1}$ would be obtained. These values are certainly minima, as the effusive component, which is definitely present and which accounts for CO_2 concentrations of 87% at a depth of 80 cm, was not taken into consideration. As the flux values reported by Baubron *et al.*¹ are an order of magnitude lower than the minima previously calculated, the reliability of their method is called into question.

Finally, we do not believe the deep source of the beach fumaroles can be distinguished from that of the crater fumaroles on the basis of the CO_2/He ratio alone. Our data indicate that this ratio ranges between 1×10^5 and 5×10^5 in the crater area^{4,8,10}. The CO_2/He ratio values of the beach fumaroles (1×10^5 – 3×10^5) (refs 1, 2) fall in the same range. In addition, the $^3\text{He}/^4\text{He}$ isotope ratios of both beach and crater fumaroles have similar values¹¹ and show synchronous time variations.

B. BADALAMENTI
S. GURRIERI

Istituto di Geochimica dei Fluidi CNR,
P.M. NUCCIO
M. VELENZA
Istituto di Mineralogia Petrografia e
Geochimica,
Via Archirafi,
36 Palermo, Italy

BAUBRON *ET AL.* REPLY — Our main demonstration — that substantial gas release occurs diffusively from the slopes of the Fossa cone, in addition to fumarolic discharge — is supported by the results of Badalamenti *et al.* Our 1988 survey (soil, well, fumarole gases) was made not only on the cone but also on more than 150 sites on the main geological structures within the caldera (2.6 km^2). Our flux estimate (30 tons CO_2 per day) was strictly for soil degassing from the cone, excluding the fumarolic fields¹. However, a computation error caused an erroneous estimate, subsequently corrected¹², resulting in a value of 50 tons CO_2 per day.

Flux measurements are made with a static (gas accumulation) method¹², which gives a mean daily flow (see Fig. 2), and guarantees certain conditions such as no disturbance of soil permeability, a representative area (0.5 or 1 m^2), integrated measurements (6–24 h) and good accuracy ($\pm 5\%$). To our knowledge, this method differs from that used by Badalamenti *et al.*⁵, who measured flux through 1 cm^2 , but we have been unable to refer to the SIMP special volume. We would appreciate the chance to compare our methods.

NATURE • VOL 350 • 7 MARCH 1991

Fetal clues to handedness

SIR — The interesting ultrasound findings of Hepper, Shahidullah and White, reported in Scientific Correspondence (*Nature* 347, 431; 1990), indicate that almost 95% of fetuses *in utero* show a bias towards sucking the thumb of the right hand. But whether this observation “supports genetic explanations of handedness” remains an open question.

The authors speculate that advanced right-sided neuromuscular development might be a trigger. This idea is presumably an extrapolation from the hypothesis that the advanced development of upper limbs compared with lower limbs in human embryos may be due to the development-promoting effects of the higher oxygen tension of the blood going to the upper limbs. Using the same reasoning, one can propose that as the right upper limb generally receives blood through a vessel that arises more proximally from the aorta than the supply to the left upper limb, development of the right might be sufficiently advanced over that of the left to promote early preference of the right thumb.

If the resolution of ultrasound imaging is adequate, it would be interesting to compare the dimensions of fetal right and left upper limbs *in utero* to see if there is a detectable difference in size that might correspond with this trophic hypothesis and its implications for handedness. This procedure might also throw some light on why a few fetuses show a preference for the left thumb. With luck, an occasional example of *situs inversus* (reversal of the normal pattern of asymmetries) might be encountered and provide an interesting test of the hypothesis, which would now predict a left-bias.

But all this sounds more like an epigenetic, rather than a genetic, explanation of handedness; it becomes a consequence of earlier asymmetries. A genetic explanation of handedness, if indeed one is required, may have to be sought at much earlier stages when asymmetries first arise in the embryonic disk.

RICHARD DRYDEN

30 Oakland Drive, Dawlish,
Devon EX7 9RN, UK

In 1989 and 1990, we carried out two more one-month programmes^{13,14}, confirming a steady distribution and high level of soil gas anomalies. CO_2 fluxes were comparable in 1988 and 1990 and lower by a factor of 2 in 1989, that is 115 tons per day for the caldera floor in 1990. Our results show that intermittent but intensive field campaigns can provide reliable data sets on degassing processes and budgets, although we agree with Badalamenti *et al.* that long-term monitoring requires continuous gas survey. We monitored CO_2 , He and Rn in a water well at the base of the cone continuously for six months in 1988 and 1989, during which we recorded variations synchronous with those at the crater¹⁵. This supports our claim¹ that

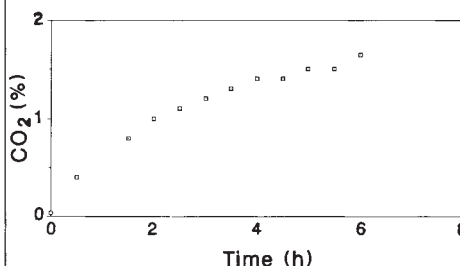


FIG. 2 Example of CO_2 flux measurement (static method) from the foot of the cone. Flux is computed from the temporal increase of the CO_2 concentration in a box. C, gas concentration in the box; C_e , limit gas concentration in the box (concentration of the gas entering into the box); C_a , atmospheric concentration of the gas; Q_e , exhaust flow of gas (equivalent to the input); T, time. The experimental data fit the theoretical curve ($C = C_e + (C_a - C_e)e^{-Q_e T/V}$). The specific flow measurements are extrapolated to the whole surface using soil atmosphere concentrations and the direct relationship = $0.5 + 0.07 (\text{CO}_2\%)$, which has remained steady over the past 3 years.

such distal gas leaks are suitable for geochemical monitoring.

Finally, our data and many others attest persistent and steady He depletion in the beach compared to crater fumaroles. This alone is not sufficient to infer a distinct origin for the fluids, but it does set important constraints on possible genetic models.

J. C. BAUBRON

BRGM, Service Géologique
National/GCH,
F-45060 Orléans Cedex 2, France

P. ALLARD

CFR, CNRS-CEA, CENS,
F-91191 Gif sur Yvette Cedex,
France

J.-P. TOUTAIN

Observatoire Volcanologique du Piton de
la Fournaise,
F-97418 Plaine des Cafres, France

1. Baubron, J.-C., Allard, P. & Toutain, J.-P., *Nature* **344**, 51–53 (1990).
2. Carapezza, M., Nuccio, P. M. & Valenza, M. *Bull. Volcan.* **44**, 547–563 (1981).
3. Gurrieri, S. & Valenza, M. *SIMP Carapezza Spec.*, Vol. 1151–1158 (1988).
4. *Rapporto GNV al Ministero per la Protezione Civile sull'intervento, straordinario a Vulcano (estate 1988)*.
5. Badalamenti, B., Falsaperla, S., Neri, G., Nuccio, P. M. & Valenza, M. *Boll. GNV* 37–47 (1986).
6. Badalamenti, B., Gurrieri, S., Hauser, S., Parelo, F. & Valenza, M. *SIMP Carapezza Spec.* Vol. 893–899 (1988).
7. Badalamenti, B., *et al.* *Boll. GNV* 519–530 (1986).
8. Badalamenti, B., Gurrieri, S., Hauser, S., Tunani, F. & Valenza, M. *SIMP* 39(2) 367–378 (1984).
9. Bertrami, R., Antrodocchia, R. & Luzi, C. *SIMP* 43 153–161 (1984).
10. Cannata, S., Hauser, S., Parelo, F. & Valenza, M. *Soc. Ital. Miner. Petrol.* **43**, 153–161 (1988).
11. Sano, Y., Wakita, H., Italiano, F. & Nuccio, P. M. *Geophys. Lett.* **16**, 511–514 (1989).
12. Tonani, F. *Gruppo Nazionale per la Vulcanologia. GNV-CNR meetings* (Roma 1990).
13. Baubron, J.-C., *BRGM Rep. R30110* (1989).
14. Baubron, J.-C., Miele, G. *BRGM Rep.* (in the press).
15. Toutain, J.-P., *et al.* *Bull. Volcan.* (in the press).
16. Baubron, J.-C., *et al.* *Bull. Geol. Soc. London* (in the press).

Delayed density-dependence

SIR — Population ecologists have frequently failed to detect significant density-dependence, the pivotal prediction of the theory of population dynamics, in natural populations of animals. Various statistical techniques for demonstrating density-dependence have been described and tested (for example see refs 1, 2), but many difficulties still remain^{3,4}.

closely for moths, but we found less delayed density-dependence than expected by chance for aphids (the model is not very appropriate for aphids with several overlapping generations per year).

We are not suggesting that the delayed density-dependence demonstrated by Turchin is not real or important, but we do suggest that the high incidence detected by him is due to the selection of forest pest species with a tendency towards outbreak dynamics. More generally, the kind of data

Our results strongly suggest that delayed density-dependence is rare in these insects and, by implication, in most insect populations. Although delayed density-dependence is uncommon in British moths and aphids, we do find non-delayed density-dependence to be common (manuscript in preparation).

ILKKA HANSKI

Department of Zoology,
University of Helsinki,
P. Rautatiekatu 13,
SF-00100 Helsinki, Finland

IAN WOJWOD

Department of Entomology &
Nematology,
Institute of Arable Crops Research,
Rothamsted Experimental Station,
Harpenden, Hertfordshire AL5 2JQ, UK

DELAYED DENSITY-DEPENDENCE IN INSECT POPULATIONS

Taxon	Number of species	series	Significance level	Observed fraction	χ^2	P
Moths	357	4,306	0.05	0.045	1.13	0.287
			0.01	0.014	2.84	0.092
Aphids	92	1,405	0.05	0.022	15.62	<0.001
			0.01	0.004	3.22	0.073

Lengths of the time-series are from 10 to 24 years.

Turchin has pointed out⁵ that most current techniques are unsuitable for detecting delayed density-dependence. Therefore, if population regulation with time lags is common, the overall level of density-dependence may be underestimated.

Turchin demonstrated this problem with an analysis of 14 sets of time-series data on forest insects, eight of which showed clear evidence of delayed density-dependence. We have analysed a large set of time-series data on annual abundances of moths and aphids using the light-and suction-trap networks of the Rothamsted insect survey⁸ (our full analysis will be published elsewhere). We have looked at the occurrence of delayed density-dependence using Turchin's method⁵, and present our results in the table.

We counted the fractions of time-series that showed significant delayed density-dependence at 5 and 1% levels. (By chance, one would expect that 5 and 1% of the time-series would show a delayed effect at these levels.) The observed fractions of significant delayed effects match the expected values

typically analysed by population ecologists are biased towards the most abundant and outbreak species, which means that the current population-dynamic conclusions based on these data should be viewed with caution⁴. Our data consist of 'ordinary' insect populations, without any bias with respect to their abundance or dynamics.

DNA damage in the Antarctic

SIR — The Antarctic ozone hole represents a cycle of annual springtime depletion of stratospheric ozone over the continent and surrounding ocean areas, with consequent increases in levels of ultraviolet-B radiation (UV-B, 280–320 nm)¹. The ecological impact of this increased radiation is of particular concern for Antarctic marine communities, which depend on the productivity of phytoplankton at the base of short food chains. There has been much speculation about ecological effects based on *a priori* assumptions about unique photobiological properties of Antarctic organisms. A common misconception is that because these organisms inhabit an area of relatively low ambient ultraviolet intensity, Antarctic species may be more sensitive to UV-B exposure than their temperate or tropical counterparts². But Antarctic species are derived from temperate and tropical stocks and should therefore retain a capacity for biochemical protection from ultraviolet exposure³ and for repair of radiation-induced damage.

During 1987 and 1988 we carried out one of the first on-site evaluations of ultraviolet exposure and photobiological responses of nine phytoplankton species at Anvers Island in the Antarctic⁴. We discovered that biologically effective UV-B can be detected as deep as 20 m during springtime ozone depletion⁵ and noted that the morphology of the cells influenced the amount of initial DNA damage⁴. The nine diatom species, which spanned nearly three orders of magnitude in

cell volume with considerable variation in surface area: volume ratios, exhibited a nearly 100-fold range in the total number of photoproducts induced. All species investigated showed evidence of both photoreactivation and excision repair at ambient Antarctic temperatures, but with considerable species variation. *Chaetoceros neglectus*, for example, removed 31% of cyclobutane dimers in 6 h but did not repair any (6–4) photoproducts; *Corethron cryophilum* repaired 46% of the dimers induced and 94% of the (6–4) photoproducts⁴.

The ozone hole has existed for more than ten years. Obvious catastrophic events have not been noted. Our results indicate that Antarctic phytoplankton have diverse capacities for sustaining and repairing UV-induced damage to DNA and as a group are not defenceless against this environmental stress. Cumulative biological effects of the ozone hole are apparently subtle alterations in species interactions, and it may take longer than a decade for these perturbations to have an observable effect. The UV photobiology of individual species needs to be a key focus in future research efforts.

SENEB KARENTZ

JAMES E. CLEAVER

DAVID L. MITCHELL

Laboratory of Radiobiology and
Environmental Health,
University of California,
San Francisco, California 94143, USA

1. Frederick, J. E. & Snell, H. E. *Science* **241**, 438 (1988).
2. Calkins, J. & Thordarson, T. *Nature* **283**, 563 (1980).
3. Karentz, D. et al. *Marine Biol.* (in the press).
4. Karentz, D., Cleaver, J. E. & Mitchell, D. L. *J. Phycol.* (in the press).
5. Karentz, D. & Lutze, L. H. *Limnol. Oceanogr.* **35**, 549–561 (1990).

Scientific Correspondence

SCIENTIFIC Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words and 5 citations, and to contributions using simple language.

If new results are being described, priority is generally given to communications that do not describe work in which the author is involved. Authors of contributions of this nature should explain in a covering letter why theirs has a particular claim on *Nature's* space. Contributions may be sent to referees and, in the case of matters arising from material published in *Nature*, are sent to the author of that article for comment. A more detailed guide to authors is available from Washington or London. □

Footnote to history

Brian Pippard

Too Hot to Handle: The Race for Cold Fusion. By Frank Close. Allen/Princeton University Press: 1990. Pp.376. £14.99, \$24.95.

THE announcement from the University of Utah on 23 March 1989, by Martin Fleischmann and Stanley Pons, that they had succeeded in fusing deuterium nuclei by electrolysis, using a palladium cathode, was greeted in different laboratories with scepticism, delight or foreboding — and by some with dismay because, yet again, the traditional process of scientific publication had been by-passed. The sceptics were solid-state physicists who could not see why embedding in a palladium lattice helps deuterons to get close enough together to fuse, or nuclear physicists who marvelled that with their declared fusion rate Pons and Fleischmann were not already dead of neutron irradiation. For the normally preferred fusion process for two deuterons is not into ${}^4\text{He}$ but into ${}^3\text{He}$ plus a neutron; hence the alarm of those who foresaw the lowering of one of the barriers to hydrogen bomb proliferation. It is not, I think, generally known that it was claimed the work had been proceeding for five years, or that it was initiated on the strength of a calculation that no moderately intelligent graduate student, let alone experienced electrochemists, ought to take seriously. Nor did it become clear until later how little there was to show for five years of work on a project of exceptional importance, one that was to be presented in the press as promising unlimited power at a negligible cost.

The lure was irresistible. Crash programmes were started all over the world to investigate the astonishing claim, and by mid-April, despite the intervening Easter holiday, three separate teams had given press releases of confirmatory evidence for nuclear fusion. Texas found heat, Georgia neutrons and Seattle tritium — but within a few weeks all three traced sources of error in their experiments and withdrew their confirmations. Each, like Pons and Fleischmann, had fallen into the trap of taking one isolated effect as proof, when only the simultaneous detection of heat, neutrons and tritium (or ${}^4\text{He}$) in roughly correct proportion can inspire confidence. But it is to the credit of all three groups that they bravely acknowledged their

mistakes as soon as they became aware of them. Pons and Fleischmann, however, according to Frank Close's account, were thrust into hasty publication by various influences, especially from patent lawyers and senior officers of the university whose immediate concern lay more with financial benefits than with what must at the time have looked like excessive academic scrupulosity. Having gone public and instigated an



A flash in the pan — B. Stanley Pons (left) and Martin Fleischmann display a model of the flask in which they said they created sustained nuclear fusion reactions.

immense media sensation, they have never retracted their claim but, faced with awkward criticism from experts, have replied evasively or even found reason to modify the data originally presented.

But if the University of Utah and a few of its chemists give capital opportunity for gibes at the unacceptable face of science, almost everyone else who became seriously involved has displayed an admirable degree of intellectual probity. However fervently they hoped to confirm cold fusion, they adopted traditional scientific standards of scepticism about their own work, and performed all the

cross-checks which Pons and Fleischmann had left undone during those hidden five years when they consulted none of their colleagues, and played around with instrumentation of which they understood all too little. The message that needs to be proclaimed loudly is that, however misguided or wrong a few individual scientists may be, the institution of science is robust. Small mistakes by their triviality may long survive undetected in a literature that is no longer consulted, but mistakes over major issues are picked up quickly — in this case, very quickly indeed. Within a month or two the initial euphoria had faded; less than two years later the episode is hardly more than a footnote to history.

Yet although a few reputations have been tarnished, others have been enhanced. The

scrupulous and painstaking investigations at Harwell and elsewhere, the balanced reporting by some leading science journalists, the tight-rope acts of editors whose professional standards proved stronger than the desire for a scoop, all these must be applauded as exemplary illustrations of what good science involves. And to these must be added, as deserving high praise, the care and industry Frank Close has shown in recording, at first hand wherever possible, the whole story from the beginning. We possess too few detailed case-histories of science, and this is a very welcome addition, even if its value lies more in the compilation than in the literary quality. For Close seems to have been overwhelmed by the demands of telling a tale that is both fast-moving and disjointed. So many points of view needed to be presented, so many simultaneous happenings faithfully recorded, that he has sought to keep the historical perspective clear in the reader's mind by repetitions which sometimes become irritating. His best passages, however, have a racy vigour; as in good thrillers, one can

hardly wait to see what they will get up to next, and as in good thrillers, what they get up to is frequently worse than expected. How gratifying that right eventually triumphs.

The book should be read as an exemplary tale by all who are concerned about the conflicting demands of scientific integrity, personal ambition and public interest. □

Sir Brian Pippard is in the Department of Physics, University of Cambridge, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK.

Quantum leap?

A. Rupert Hall

Science, Optics and Music in Medieval and Early Modern Thought. By A. C. Crombie. *The Hambledon Press*: 1990. Pp. 474. £37.50, \$60.

ONE of the most disputed issues in the history of early modern science is that of continuity. Was there a quantum leap in the seventeenth century, and if so, what was its base? Alistair Crombie was a prominent pioneer of the view, now widely accepted, that there was such a leap forward ("the confident establishment of the 'new sort of philosophizing' in the 'physico-mathematical experimental learning' of the seventeenth century"), and that the Middle Ages had made highly important contributions to the base from which the quantum leap took place. Though mediaeval academics were stronger in philosophizing than in experimental or mathematical achievements they enriched the ancient Greek idea of nature, and set the potential of this enrichment in a stronger, more fertile context of society and technology. To these elements the Renaissance, drawing afresh upon the founts of European civilization in antiquity, added new ones such as platonism, hermeticism and atomism; to these surely geometry should be added.

Choosing 18 papers (and some annexed documents) for revision and reprinting in facsimile, Crombie has devoted nearly half to this earlier and major theme in a scholarly career both long and distinguished. His investigation of mediaeval and Renaissance science during the last 20 years is notably broader, more complex and more balanced than it was in his 1953 monograph *Robert Grosseteste and the Origins of Modern Experimental Science* (Clarendon Press, Oxford) in which the 'Crombie thesis' was first expressed with great learning and great force. Such a paper as that of 1980 on Science and Arts in the Renaissance admirably adds to Crombie's earlier studies of mediaeval conceptual science an emphasis upon the practical and mathematical accomplishments of the Renaissance.

In the last 20 years, too, Crombie has developed a new interest in major figures of the second phase of the scientific revolution; Galileo, Kepler, Mersenne, Descartes. Two long papers — the second a translation of Kepler's chapters on The Manner of Vision — deal with the context and nature of Kepler's treatment (1603) of the eye's formation of a retinal image, "the first thorough and successful solution of an important physiological problem to be made in modern times." Kepler's, and later Descartes', analyses of vision also had important technical implications for the design of lens-systems. The former of these two papers, which might with advantage have been placed second

here, very impressively represents Crombie's abilities in massive original investigation.

The remaining papers are all connected with the author's works which have not yet seen print, one on Galileo, the other on 'styles' of scientific thought. A pervasive theme, found also in the papers on vision, is "the power of those commitments to quantification, demonstration, and the hypothetical model [that were] brought into physiology by mathematics" in the seventeenth century — and one might add, into aesthetics also. The example is Mersenne's investigation of acoustics/music, where Galileo also figures in his own right, and his father's:

We may recognize in Vincenzo Galilei's insistence on both the complexity and on the discoverable regularities of auditory experience something of Galileo's approach to natural science, and certainly a family likeness in the polemical aggressiveness to be made famous by his son.

A fundamental issue in this connection was the distinction between primary and secondary qualities drawn from physicists by Galileo, for philosophers by Locke; this of course defines the domain of natural science as "the art of the possible". A very interesting paper of 1964 examines the junction in the seventeenth century between mechanistic philosophy and the ocular demonstrations of the anatomists: another facet of a problem continuing since Galen.

Reading or re-reading these fruits of 40 years, one finds that they present one conspectus of the evolution of the historiography of science over more than a generation. Be it said in his praise, Crombie is largely consistent with himself and has found little to rewrite. But he is occasionally ambivalent; can a disciple of Alexandre Koyr (as he ardently proclaims himself) also affirm "how misleading it would be for historians to separate the technical science of a period from its ambience of larger intellectual commitments"? Koyr did not mingle intellect with technique. In the present collection, to be followed by another, there are inevitably some repetitions and the author expects his reader to be almost as learned as himself. To unfold an allusion to "the validity in physics, in any inquiry for physical causes, of the axiomatic theory of truly scientific apodeictic demonstration embodied in somewhat different forms in Aristotle's logic and Euclid's geometry" a treatise is required. Yet, with much mental agility as well as full deployment of the heavy artillery of scholarship, Crombie has won his own considerable victories in a field dominated by German and North American scholarship. □

A. Rupert Hall is at 14 Ball Lane, Tackley, Oxfordshire OX5 3AG, UK.

Spring Books

The next review supplement to appear in *Nature* will be Spring Books to be published on 11 April.

Just words

J. R. Ravetz

The Rhetoric of Science. By Alan G. Gross. *Harvard University Press*: 1990. Pp. 248. £23.95, \$29.95.

It is quite some time now that working scientists have been enduring a bad press from the scholars who analyse their activity. Starting explicitly with P. Feyerabend's *Against Method* (New Left Books, London) of 1975, the tide has been running away from philosophical explanations of how science gets it right, to social-science analyses of the practices of science which variously deconstruct, demystify or debunk. Those who try to defend the old verities of science, in the name of 'realism', are an embattled minority.

Someone writing on the 'rhetoric' of science might be expected to belong to the pack of critics, since 'rhetoric' has come to stand for mere persuasion, in contrast to logic and scientific method which yield truth. Even more than a sociologist, a rhetorician has a professional motivation for reducing science to the same lowly status to which it had been assigned. Because (as all philosophers have agreed) the arguments of empirical science are not formally valid in the logical sense, all it needs is an all-or-nothing pattern of logic to lead to the conclusion that science is no better than poetry or propaganda.

On the other hand, because recent writings on theory of rhetoric have reclaimed it as a genuine mode of discourse, perhaps a rhetorician might be more generous than an anthropologist, and demonstrate at least an intersubjective validity in scientific discourse. The issue at stake here is not mere academic infighting, but the long-term influence on how natural science is perceived, by the general public and also by the next generation of teachers and their students.

The author of *The Rhetoric of Science* seems at first to incline to the more generous interpretation. In his first interpretative essay, on taxonomic language (is evolutionary theory inconsistent with species as natural kinds?), he envisages a complementarity between rational and rhetorical reconstruction. The one is subsumed under the cognitive/technical 'interest' (in Habermas' sense), and the other under the practical 'interest', relating to various forms of life. But when he confronts the philosophical problems inherent in such a duality, he comes out in favour of Habermas' third 'interest', which is emancipatory by its demystification of the objectivity of science.

His subjects for analysis are widely drawn, ranging from the discovery of DNA (and James Watson's retrospective autobiographical fable about it), the reception of the Copernican theory (with a good discussion of 'rational conversion'), an analysis of

'rational conversion'), an analysis of Isaac Newton's rhetorical devices in early and late writings on optics, and the origins of Darwin's *Origin*. These draw on established historical sources, and I doubt that an historian of science, of whatever philosophical persuasion, would disagree with them. There are also essays on scientific practice which seem not to entail any radical conclusions about scientific knowledge.

But in the epilogue, on 'Reference without Reality', he picks up the theme again, and seems nearly to make scientific agreement and disagreement a matter of taste. Quoting D. Davidson, he reduces Kuhn's paradigm-conflicts to how 'we' choose to phrase our differences. Similarly, he quotes N. Goodman on truth as being "like intelligence, perhaps just what the tests test." I agree with his statement (near the end) that the sciences create bodies of knowledge so persuasive as to seem unrhetorical, but to me it seems capricious to imagine this persuasion being accomplished more by rhetorical devices than by argument based ultimately (with however many lacunae and imperfections) on a controlled and shared experience of an external world. □

J. R. Ravetz is at The Research Methods Consultancy Limited, Quality Assurance for Scientific Information, 13 Temple Gardens, London NW11 0LP, UK.

Man of courage

Nevill Mott

Kapitza in Cambridge and Moscow.

Edited by J. W. Boag, P. E. Rubinin and D. Schoenberg. Elsevier. 1990. Pp.429. Hbk \$77, Dfl 150; pbk \$38 50, Dfl 75.

SINCE his death in 1984 parts of the story of Peter Kapitza have been told in several publications. *Kapitza in Cambridge and Moscow* contains his letters and has a biographical introduction in which David Schoenberg tells the story again. Born in 1884 at Kronstadt near St Petersburg, he was a student at St Petersburg Polytechnic during most of the war of 1914–1918 and married and fathered a son. But in the terrible conditions after the revolution and civil war, first his father died, then his son and finally his wife, shortly after giving birth to his daughter who died too. Naturally enough, Kapitza was slow to recover, but in 1921 Abraham Joffe, the leader of Russian physics at that time, with the support of Lenin organized a trip abroad "for renewing scientific relations with other countries." He included Kapitza in the party, who in England by sheer force of character persuaded Rutherford to take him on in the Cavendish, where he stayed for 13 years. He achieved great success in research, a fellowship of the Royal Society (though remaining a Soviet citizen) and a fellowship in Trinity College. Included in the

book are translations of his letters to his mother at that time. One quotation will suffice, "Today the Crocodile [Rutherford] summoned me twice to discuss my paper. It will be published in the Proceedings of the Royal Society, which is the greatest honour a piece of research can achieve here. It is only now that I have felt my strength. Success gives me wings and I am carried along by my work." Also, he married again.

I knew Kapitza well at that time. In 1934 we both went to the Soviet Union, as he often did. For me the occasion was a conference to celebrate the hundredth anniversary of the birth of Mendeliev. Kapitza had omitted to obtain an assurance from the Soviet embassy that he could return to England, and found that he could not. Letters to Rutherford, and to his wife, Anna, who remained in England for two years, form a major part of the book. At first he was miserable. Colleagues, not knowing if he was in favour or not, shunned him. But later his situation improved, and he showed his sincere belief in the socialist creed, and his desire to help the explosive growth of science in the socialist motherland.

Perhaps most interesting are the 'Letters to the Kremlin', to Molotov, Stalin, Krushchev, in which he showed amazing courage. He secured the release of L. D. Landau, Russia's most brilliant young theorist, after one year of a prison regime which could soon have killed him. He complains how hard it is to run a research institute on the basis of a detailed plan for the year ahead, so that it is impossible to obtain supplies to fill an unexpected need. He writes to Stalin that the

Academy of Sciences is full of third-rate old men and that the only way to change this is for the Soviet populace to show the same pride in their science as they do in their ballet. To Keldysh, President of their Academy of Sciences he writes as late as 1972 how intolerable it is that his travel arrangements are always interfered with, for instance that his permission to attend the Rutherford Centenary Celebrations at the Royal Society was withdrawn at the last moment because of the expulsion by Britain of members of the Soviet trades delegation.

Curiously, there are letters to Max Born after he left Germany in 1933 and to Niels Bohr after he escaped from Denmark in 1943, offering them permanent positions in Moscow. To Born he writes, "please consider our 1/6 of the world as a possible place for you to settle". After describing some of the hardships, he writes, "After all, the Bolsheviks are angels compared to your Nazis, and

what is much more they have a true case to fight for . . . I agree with you they are the only ones who keep the right line, and indeed a winning one."

Compared to the Nazis, yes — but in any absolute sense, could he have thought so?

There are some curious letters about a particular scheme of his for destroying missiles in flight by intense electromagnetic radiation. This was not pursued.

At the end of the war he seems to have been in charge from the scientific side of the Russian project to build an atomic bomb, and the famous letter to Stalin is reproduced, saying how impossible it was to work with



Kapitza — Fellow of the Royal Society and member of the Soviet Academy of Sciences.

Lavrenti Beria, who seems to have been their General Groves. "You don't understand science", says Kapitza. "You don't understand people", says Beria. After this, maybe ostensibly for other reasons, Kapitza was removed from the project, though not from his membership of the Academy with its salary. He continued to write to Stalin.

After Stalin's death and the execution of Beria, Kapitza came back into favour. The book ends with some letters to Andropov about Sakharov and Orlov, referring to Lenin's respect for scientists.

To anyone who knew him, *Kapitza in Cambridge and Moscow* brings back the memory of a very great man, lovable and human, and of his life under a tyranny of a kind that has not yet disappeared from the world. □

Sir Nevill Mott is at 63 Mount Pleasant, Milton Keynes MK17 8JX, UK.

A latter day green Penguin

John A. Campbell

Hermes and the Golden Thinking Machine. By Alexander Tzonis. MIT Press. 1990. Pp. 284. \$19.95, £14.95.

ARTIFICIAL intelligence (AI) has something for everybody: practitioners, scientific observers from other subjects, people with high and low literary-artistic tastes, and ordinary citizens. It is therefore surprising that there have been rather few tries at writing the Great AI Novel. Most of the existing efforts are squarely in the science-fiction tradition. Some are fun, but even these rely on vague or over-imaginative impressions of what real AI research is about, and hence what are its likely consequences in possible future worlds. Perhaps the secret of eventual success is to be indirect and allusive, not to tackle AI head-on or to put it deliberately in the centre of the stage, and not to take the enterprise too seriously. These conditions are entirely true of my personal favourite in the not-quite-AI class: *The Tin Men*, by Michael Frayn.

Superficially *Hermes and the Golden Thinking Machine* is an AI novel, with a criminal plot. Turned inside out, it looks more like a whodunit which would have been perfectly at home in the Penguin green-covered paperback crime series of the 1940s, with intellectual additions. In terms of AI, it gets a high rating for allusions, an adequate passing grade for suitable lack of seriousness, but only a 'pass if pressed' mark for confrontation. Tackle AI it certainly does, but it brings to mind the story of the Welsh full-back defending himself for losing a rugby game in the last minute by failing to stop one Fotherington-Smith in the English team from scoring a try. "I started to tackle him", he said, "But then Fotherington went one way, Smith went the other, and I tackled the bloody hyphen".

The idea that the main feature of communication (non-communication?) is the encrypting and decrypting of messages is currently fashionable in literary circles where deconstruction is a key word; indeed, some university departments have made an entire industry out of it. In principle depart-

ments of archaeology are ripe for industrial takeovers, because decryption of real or involuntary messages from extinct societies must be an even more distinguished branch of the fashion. Hermes, the book's hero, makes the author's argument for a merger by embodying it: he is a professor of archaeology and AI, who operates in both modes at once.

The idea is fine and deserves development, particularly if combined with some of the most recent preoccupations of AI research (such as the structure and use of explanation, and the aspects of knowledge representation that are needed to support it) and with insights from philosophers like R. G. Collingwood who have thought about it without the constraints of fashion. What we get instead is an undeveloped version: raw, but the taste is still agreeably exotic.

This is the Fotherington of the book. The Smith is the treatment of technical examples of AI that turn up from time to time — or, at least, introductory textbook examples that were fresh in the mid-1970s — to add a little verisimilitude. A reader does not have to be an AI specialist to suspect that Fotherington and Smith are so different that they can scarcely share the same hyphen. This is unfortunate, because a more up-to-date sample of AI would have given the means for making the two fit together much more convincingly.

But too much highbrow analysis is probably unfair to the author. The book is best read as a latter-day green Penguin. In style and content it would have been a respectable

member of that series. The plot is a straightforward romp through modern Greece with a dash of New England as various ambiguous characters strike structuralist and even post-structuralist attitudes and bump each other off in pursuit of the Golden Thinking Machine, a valuable archaeological object. The author is generous with his material, and neglects the experienced crime writer's principle of economy: never introduce any novel piece of data without exploiting it later. For example, the obvious potential for the Machine to prop up the philosophical and message-encryption end of the plot is given only perfunctory attention. Nevertheless, the book is a good enough combined criminal and technical read that one should not be too pedantic about rules of genre.

Despite publication by MIT Press, the main technical interest is probably not AI. As a pure AI frolic, it moves in heavy boots. A more sporting question is whether or not it was written to send up the structuralists and their still more modern friends. It stands either that interpretation or viewing as an accidental send-up by an author, using old AI ideas as tools, who has never deconstructed a text in his life. Either way, it works. A book that can be treated as both a detective story and a reflection of what some of our faculty of arts colleagues are up to is well worth consideration for one's next free long weekend. □

John A. Campbell is in the Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK.

Long-term study

T.C. Whitmore

Four Neotropical Rain Forests. Edited by Alwyn H. Gentry. Yale University Press. 1990. Pp. 627. £42.50, \$66.50.

THIRTY-four scientists have contributed to *Four Neotropical Rain Forests*. It describes the four places in the New World tropical rain forests which have seen the greatest concentration of research by US biologists of many different disciplines. All four sites have therefore become very familiar in name to tropical biologists. The oldest is Barro Colorado Island (BCI), Panama, where work began 65 years ago when the island was created during the construction of the Panama canal. Following in age sequence are La Selva, Costa Rica, where research began in 1969, Manu in the Peruvian Amazon (1969/70) and the Minimum Critical Size of Ecosystems (MCSE) project area 80 km north of Manaus in Brazil (1979). First the sites themselves are described, then follow sections on floristics, birds, mammals, reptiles and amphibians and forest dynamics. In every section there is a chapter for each site and a concluding chapter of comparison and synthesis. Unfortunately, for MCSE floristics

a different site 80 km away is described. For the subject covered, the book gives a summary of the state of knowledge, which is inevitably uneven, and this affects the synthesis chapters.

Much stronger editorial intervention would have been useful, to cut out, for example, considerable repetition of site description and climate, and to provide a complete set of maps to show the outsider where the sites are and what they are like. The suggestion is made in many chapters that species richness is related to productivity and that in turn to soil fertility, but no data are presented and few references given to work to investigate the validity of such a speculation.

The book is best regarded as an introduction to these four much mentioned sites and to the work done so far on the subjects chosen for coverage. For these it provides a useful overview and a way into the scattered research literature, which however needs to be consulted for more critical discussion and evaluation of the data.

The book strongly demonstrates the value of long-term studies, and the important synergy sometimes achieved where different facies of highly complex rain forest ecosystems are studied at the same location. □

T. C. Whitmore is in the Department of Geography, Cambridge University, Downing Place, Cambridge CB2 3EN, UK.

New in paperback

The Causes of Evolution by J. B. S. Haldane — a classic of evolutionary theory — has just been republished by Princeton University Press with a new afterword by Egbert G. Leigh Jr. Price is \$7.95, £4.

From Mandarin comes *The Driving Force: Food in Evolution and the Future* by Michael Crawford and David Marsh who challenge the idea of survival of the fittest and argue that evolution is a process of cause and effect with nutrition playing a determinant role. Price is £5.99.

The evolution of mating preferences and the paradox of the lek

Mark Kirkpatrick & Michael J. Ryan

Why do females prefer elaborate male mating displays in species where they receive little more from males than their sperm? Here we review three hypotheses for the evolution of mating preferences: direct selection, the runaway process and the parasite mechanism. There is growing support for direct selection, in which preferences evolve because of their direct effects on female fitness rather than the genetic effects on offspring resulting from mate choice.

THE evolution of the exaggerated mating displays made by animals that meet in groups, or leks, during the breeding season is one of the most enduring problems in evolutionary biology. In animals as diverse as the birds of paradise, guppies, frogs and insects, males congregate in leks to display to females (Fig. 1). Each female chooses from among the assembled males, and seems to receive little more from her mate than his sperm. Females often show strong unanimity in their choice, and so a few males acquire most of the matings. The successful males are typically those with the most extreme plumage, vocalizations and displays. Consequently these secondary sexual characters have evolved to such extremes that they decrease male survival (Fig. 2). Why should females have evolved such strong preferences when they seem to receive no tangible benefit from their choice? This puzzle has been called the 'paradox of the lek'.

The status of this problem has recently been transformed by two developments. First, detailed mathematical models have allowed the identification of a variety of evolutionary forces that could cause the evolution of preferences for exaggerated male displays (Table 1). This situation contrasts strikingly with that of just a decade ago, when only two mechanisms were discussed and even those were not fully understood. Second, there is now a growing body of empirical research attempting to determine which of these mechanisms are actually operating in natural populations. Current efforts involve quantitative field studies, experiments under controlled laboratory conditions, and phylogenetic (cross-species) comparisons.

Despite substantial efforts, the primary factors responsible for the evolution of preferences remain controversial. Most recent attention has focused on three of the possibilities listed in Table 1: direct selection on preferences, indirect selection of preferences through their interaction with parasites (the Hamilton-Zuk¹ hypothesis), and indirect selection of preferences by a 'runaway process'. In this review we synthesize our current understanding of these three hypotheses; a broader perspective is given in other recent reviews^{2,3}. We begin by describing a 'null model' as a reference for understanding the effects of different mechanisms proposed for preference evolution. We then consider the three specific hypotheses, first describing their theoretical bases and then reviewing the empirical evidence. Although we limit our discussion to these mechanisms, we emphasize that they are only three of many possibilities, and that more than one of them could operate simultaneously (see Box).

Three hypotheses for preference evolution

Before discussing specific hypotheses, it is helpful to understand what happens in a null model that has been intentionally stripped of all complications. The simplest possible situation involves the evolution of a male trait that affects both male survival and attractiveness to females, and a preference that influences how females mate but that has no other effects on female survival or fecundity⁴⁻⁶. (In the evolutionary sense, a 'preference' is any

trait that biases the probabilities that females mate with different kinds of males, and does not necessarily involve any cognitive process.)

Under the null assumptions, several genetic models reveal an evolutionary equilibrium that balances male survival against male mating success. The degree of elaboration in the male trait that produces this balance depends on the average preference in females. Consequently, there is a line or curve of evolutionary equilibria relating the average female mating preference to the average male trait in the population (Fig. 3). Once a population reaches a point on this curve, there will be no further evolutionary tendency in the female preferences. Specifically, if a preference for one type of male is established by some mechanism, then alternative preferences for males that survive better (or worse) will not be favoured^{5,6}. Although females mating with males that survive better will have sons of high viability, those offspring will have lower mating success than other males with traits that are more preferred by females. While this simple picture becomes more complicated when certain assumptions about the behaviour and underlying genetics are made⁷⁻⁹, it serves as a useful benchmark for discussing more complex situations.

The null model further shows that a genetic correlation (resulting from linkage disequilibrium) between the preference and male trait can develop because females with the most extreme preferences mate with the most extravagant males and so produce offspring that carry genes for extreme values of both the

TABLE 1 Proposed mechanisms for the evolution of mating preferences in polygynous animals

Mechanism	References
1. Direct selection of preferences	16
A. Males provide resources to females or offspring	12, 15, 17
B. Costs of searching for mates	11, 13, 14, 51
C. Selection against hybridization	52, 57
D. Males differ in sperm fertility	18
E. Pleiotropic effects of preference genes	
F. Disease or parasite transmission	
2. Indirect selection of preferences	
A. Runaway process	5, 10
B. Good genes	
(i) Host-parasite coevolutionary cycles	21, 53, 58
(ii) Unconditionally advantageous mutations	13, 17
(iii) Unconditionally deleterious mutations	21, 54, 55
C. Genetic epistasis and dominance	8, 9
D. Social system	7
E. Mutation pressure on trait	16
3. Other mechanisms	
A. Random genetic drift	5
B. Group selection	56
C. Mutation pressure on preference	16

References are to detailed theoretical studies.

Interaction of direct selection and the parasite mechanism

Female mating preferences may experience several forms of selection simultaneously. One way to assess the evolutionary potential of these mechanisms is to consider how they interact^{11,13,14}. Here we consider the interaction of direct selection (resulting, for example, from search-costs or pleiotropic effects of preference genes) and the parasite mechanism using a highly simplified quantitative genetic model.

A form of direct selection that may act on many preferences is stabilizing selection. When variation in preference genes affects female survival or fecundity, there may be an optimum value for the preference that maximizes female fitness (see Fig. 4, top). The per-generation evolutionary change in the average preference of females caused by direct stabilizing selection is approximately

$$\Delta\bar{p} = \frac{G_p^2}{\omega^2} (\theta - \bar{p}) \quad (1)$$

where $\bar{p}$ is the mean of the preference among females, θ is the preference optimum, G_p^2 is the additive genetic variance for the preference, and ω^2 is the width of the fitness function favouring the optimum preference θ (smaller values of ω^2 correspond to stronger stabilizing selection towards the optimum). An implication of equation (1) is that the force of selection tending to restore the mean preference to its optimum is proportional to the deviation of the mean from the optimum.

Now consider the effects of indirect selection generated by the parasite mechanism. Female preferences become exaggerated as a side-effect of the evolutionary increase in parasite resistance. The per-generation change in the mean preference caused by the parasite mechanism is approximately

$$\Delta\bar{p} = G_{rp}\beta_r \quad (2)$$

where G_{rp} is the additive genetic covariance between the mating preference and the parasite resistance, and β_r is the selection gradient (a measure of the strength of directional selection) acting on parasite resistance. (A more detailed model would account for the oscillating selection pressure that from the parasite hypothesis is postulated to be acting on parasite resistance.)

These equations illustrate an important qualitative difference between direct and indirect selection. With direct selection, the strength of selection depends on the mean preference in the population (equation (1)). By contrast, the strength of indirect selection produced by the parasite mechanism is independent of the mean preference (equation (2)). When both forms of selection operate, the equilibrium for the mean preference is at a point that balances these two forces:

$$\bar{p} = \theta + \frac{\omega^2 G_{rp} \beta_r}{G_p^2} \quad (3)$$

At equilibrium, the preference will lie in a vicinity of its optimum, θ , but will deviate from it by an amount proportional to the strength of the indirect selection caused by the parasite mechanism.

Two conclusions follow. First, the parasite mechanism will pull the mean preference away from its fitness optimum. By decreasing the average fitness of females, the parasite mechanism lowers the population's total reproductive output (mean fitness). Thus 'good genes' mechanisms can actually have deleterious evolutionary effects. Second, the parasite mechanism will not cause indefinite elaboration of the preference and male trait when direct selection is also present. The preference is evolutionarily constrained by direct selection to the vicinity of its optimum. The parasite mechanism will only have a substantial effect on the evolution of the preference when direct selection is relatively weak compared with the strength of indirect selection resulting from the parasite mechanism.

trait and preference^{4,10}. The genetic correlation does not itself cause evolution of the preference, but is a necessary component of indirect selection, as we will discuss below.

An important message from the null model is that neither preferences for extravagant traits that decrease male survival nor preferences for traits that enhance male survival will necessarily spread among females without the action of some additional evolutionary force. What might that be? Most attention has focused on three hypotheses: direct selection of prefer-

ences, indirect selection in a runaway process, and indirect selection by the parasite mechanism.

■ **Direct selection of preferences.** Direct selection on mating preferences arises whenever the preferences affect the survival or fecundity of females. Direct selection results from many processes (Table 1). If the females that prefer more conspicuous males, for example, spend less time searching for mates than females preferring drab males, they will be favoured by direct selection. A second and possibly important form of direct selection occurs when preferences have pleiotropic effects that affect female survival and fecundity through effects unrelated to mate choice. In fact, some form of direct selection will act on preference genes unless they have absolutely no effect on immediate fitness. The simple generalization emerging from theoretical studies is that direct selection favours preferences that increase the average fitness of females. This result occurs with models of a variety of mechanisms that produce direct selection^{3,5,11-17}, and was recently rederived by Grafen¹⁸. Once a mating preference is established by this process, it dictates the equilibrium for the male trait (Fig. 4). Although adaptive from the females' point of view, these preferences can cause the evolution of elaborate male displays that decrease male survival.

■ **Indirect selection of preferences: the runaway process.** The second and third hypotheses we discuss both involve indirect selection. This mode of evolution depends on the evolutionary exaggeration of a character that is genetically correlated to the mating preference. Evolution of that character causes the preference to be exaggerated as a side-effect. Although a genetic correlation is not required for the exaggeration of the preference or the male display trait under direct selection, it is necessary for indirect selection to operate.

Evolution of a male display trait itself can cause a preference to evolve this way, because a preference and its trait can become genetically correlated (see above). If the evolutionary exaggeration of the preference is sufficient, theory shows that an unstable runaway process can result^{5,10}. In a runaway, the male trait cannot reach an equilibrium because the force of sexual selection generated by the ever-more extreme preference accelerates more rapidly than the trait can evolve (Fig. 4). Ultimately, changes in either the genetic variation in the population or the force of viability selection acting on the male trait would bring the process to a halt. A runaway could cause substantial evolution of the trait and preference in a small number of generations. Like direct selection, a runaway will generally establish preferences that are arbitrary with respect to male survival. For a runaway to be initiated, the genetic correlation between the preference and trait must be larger than a threshold determined by the strength of viability selection acting on the male trait and the form of the mating preference⁵. Several factors, including direct selection on preferences and random genetic drift, can decrease the possibility of a runaway occurring. Whether the conditions for a runaway have ever been realized in a natural population is unknown.

The runaway process was first discussed by the theoretical population geneticist R. A. Fisher¹⁰ in 1958 and earlier. The attention that it now receives comes more from its historical importance as the first modern hypothesis for preference evolution than from any empirical support for it.

■ **Indirect selection of preferences: the parasite hypothesis.** The third hypothesis has its roots in the experience of many field naturalists. They have often observed that females in species that form leks seem to mate with the most vigorous and healthy males. This impression naturally led to the so-called 'good genes' hypotheses for preference evolution. The intuitive argument behind all good genes hypotheses is that preferences for males with genes that enhance viability are favoured by evolution. By mating with a vigorous male, a female gains an evolutionary advantage by passing those genes on to her offspring^{19,20}. A serious challenge to this logic appeared when the first null models showed that preferences for high viability males will not



FIG. 1 Clockwise from top left: Male bird of paradise, *Paradisaea raggiana*, performing a courtship display to a female (courtesy of B. Beehler). Male moth, *Creatonotos gangis*, attracting females with pheromones (courtesy of

M. Boppré). Calling male glass frog, *Centrolenella fleischmanni*. Male platyfish, *Xiphophorus variatus*, and male swordtail, *X. helleri*.

FIG. 2 Calling male túngara frog, *Physalaemus pustulosus*, being eaten by a bat, *Trachops cirrhosus* (courtesy of Merlin Tuttle, Bat Conservation International).



necessarily spread even in the ideal situation where females can correctly identify those males (see above). More complex variations of the good genes argument have been developed, however, and recent theoretical work has confirmed that these mechanisms could work under some circumstances^{17,21}.

The good genes hypothesis that has attracted the most attention is Hamilton and Zuk's parasite hypothesis¹. They noted that parasites are ubiquitous, and suggested that continuous coevolution between parasites and their hosts might drive the evolution of preferences for extreme male displays. Hamilton and Zuk postulated a correlation between the genes for resistance to parasites and the expression of the male trait. A variety of mechanisms could produce this correlation, the simplest being that more resistant males will be healthier and in better condition to grow elaborate plumage and perform strenuous displays^{21,22}. Then females with preferences for the most extreme males will also tend to mate with those males that are the most resistant to disease. This establishes a genetic correlation between the preference and resistance genes so that the evolution of greater parasite resistance also causes the evolution of more extreme preferences. The male trait will become exaggerated as a result (Fig. 4). The critical assumption that keeps the trait and preference from reaching an equilibrium as it does in the null model is that the genes responsible for resistance are assumed to be constantly changing.

Predictions, tests and inferences

Although these three hypotheses differ in several fundamental ways, no definitive support for any of them has so far been obtained for two reasons: variation in mating preferences is more difficult to quantify than it is for many other types of traits (such as male displays), and many observed preferences are consistent with several evolutionary hypotheses. Two approaches have been used to test the hypotheses: studies of single species and cross-species comparisons. Here we outline the data that would be ideal to test each hypothesis and assess the evidence that is presently available.

■ **Direct selection.** Direct selection on preferences could be demonstrated by observing a correlation between a female's mating preference and her survival or reproductive success. So far, there are no studies showing directly that variation in female preferences affects female survivorship or fecundity. But there are data that implicate direct selection.

In species that do not form leks, female preferences seem to have evolved to maximize resource quality or quantity. Abundant data show that when males provide a nest site, food or care for the young, females prefer mates who provide resources that enhance female fecundity²³⁻²⁵. But the paradox of the lek derives from female choosiness in the absence of such material resources. Direct selection on preferences in species that form leks must result from other factors.

A subtle form of direct selection may be very important. It occurs when genes that affect the female's mating preference also affect other aspects of her life, the well known phenomenon of pleiotropy^{3,26,27}. Because females use their sensory systems for other tasks besides mate choice, these systems will often be subject to natural selection for other reasons, such as foraging ability or predator detection, with the side-effect that preferences for traits that decrease male survival are likely to be established. One example comes from studies of insectivorous anolis lizards²⁸. Their visual system is exquisitely adapted to detect the motion of prey. The male 'pushup' courtship display seems to have evolved to match these sensory biases in order to attract the attention of females. Other possible examples of pleiotropy come from how selection for predator detection might result in ommatidial organization that also favours construction of courtship pillars in crabs²⁹, and how selection on body size to avoid desiccation might influence call-frequency preference in frogs³⁰.

Several mechanisms other than pleiotropy can also cause direct selection in species that form leks. All females receive sperm from their mates, and in some of these species the males' size and physical condition might betray variation in their ability to fertilize. Direct measures of female fecundity have shown that in frogs³¹ and fruitflies³², females prefer to mate with males whose phenotypes maximize fertilization success. Contagion of diseases and ectoparasites will select for female preferences to avoid transmission. This effect has been implicated in the evolution of mating preferences in several animal groups^{33,34,59}. The time, energy and risk of searching for a mate can also expose preferences to direct selection. Although female search costs have been identified in some species³⁵, there is still no evidence that variation in search costs results in selection favouring particular preferences in any species that form leks.

Phylogenetic comparisons have recently been used with two groups of animals to reject the possibility that their mating preferences originated through indirect selection caused by either the runaway or parasite mechanisms. The results suggest some form of direct selection was involved. The frog *Physalaemus pustulosus* (Fig. 2) adds chucks to its introductory whine call and females prefer calls with chucks³⁶. Only *P. pustulosus* and its sister species *P. petersi* produce chucks, and the chuck seems to have evolved in their most recent common ancestor. In the close relative *P. coloradorum*, however, females also prefer calls with chucks, despite the lack of this call component in their male's repertoire. This suggests that the preference was present in the common ancestor of all these species, and thus the preference evolved before the chuck (M. J. Ryan and A. S. Rand, personal communication) (Fig. 5). The fish genus *Xiphophorus* consists of two groups, the swordtails and the platyfish. Swordtails have a sword-like elongation of the caudal fin for which females show preferences. The sword must have evolved after the swordtail and platyfish lineages diverged

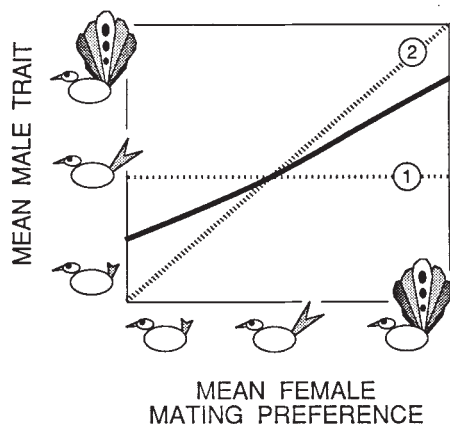


FIG. 3 Equilibria for preference and trait in the null model (heavy line). Females are imagined to have an absolute search-image for the type of male with which they would prefer to mate. The evolutionary equilibrium for a male display trait (such as tail length) represents a compromise between what maximizes survival (dashed line ①) and what maximizes mating success, which depends on prevailing female preferences (dashed line ②). (A qualitatively similar equilibrium curve results when mating preferences are open-ended rather than absolute.)

because platyfish lack swords (Fig. 1). But female platyfish show preferences for conspecific males with artificial swords³⁷. As in *Physalaemus*, the male trait in swordtails apparently evolved to match pre-existing female mating preferences (a process called sensory exploitation^{26,38}). The female preference and its corresponding male trait did not coevolve together as predicted by the indirect selection hypotheses. The data are consistent with the idea that direct selection established sensory biases with pleiotropic side-effects that influence mate choice.

■ **Indirect selection: the runaway process.** Two kinds of observations would be needed to document a runaway process: an imbalance between natural and sexual selection that was causing directional evolution of the male trait, and a genetic correlation between the preference and trait sufficient to cause the preference to evolve fast enough to maintain the imbalance. It is technically difficult to collect these data in most natural popula-

tions. Further, it is unlikely that a runaway would be observed in action because it is not expected to last for evolutionarily substantial periods of time. There also seems to be little hope in making predictions that could easily be tested by comparative methods because the direction of evolution during a runaway is unpredictable. The prospects for confirming the runaway hypothesis therefore seem to be dim.

There have been attempts to observe one of the results of a runaway process by comparing populations to see if those that have the most extreme preferences also have the most extreme male traits. This correlation has been found in guppies, where the elaboration of male colour is correlated with the degree of female colour preference among populations^{39,40}. But this correlation offers only weak support for a runaway because it can also result from other mechanisms for preference evolution^{39,40}.

■ **Indirect selection: the parasite hypothesis.** To observe directly the parasite mechanism in a given species, one would need to find a genetic correlation between the resistance genes and the female mating preference, and document the evolutionary increase in parasite resistance. Because these kinds of observations are impractical in most species that form leks, studies of single species have turned to more limited predictions. Møller³⁴ recently reviewed almost a dozen such studies. Most show that the expression of male traits is correlated with parasite load and that females prefer males with fewer parasites. In five studies that looked for heritable variation in parasite resistance, it was found. At a recent symposium⁴¹ devoted to this topic, six studies were reported in which traits and preferences were found to be related to parasite load, but in three other studies this relationship was reported to be absent.

Several workers have pointed out that many results consistent with the parasite hypothesis are not necessarily strong evidence in its favour. Consider first the data showing that females prefer mating with less parasitized males. There are many reasons why this might be so. No matter what established a preference, it is likely that healthy males will be more successful than sick males in attracting females. For example, consider a population of plants in which healthy individuals make more flowers and attract more pollinators than diseased individuals: surely this is not because the pollinators have evolved preferences for disease-resistance genes in the plants. A further complication is that preferences to avoid diseased males could be established by direct selection as well as by the parasite mechanism^{33,42}. In species with male parental care, including most birds, females who mate with healthy males are at a selective advantage because these males will be better able to help rear young than will sick males. Females that avoid contact with contagious males are also favoured by direct selection (see above). In short, data showing that females mate with less parasitized males are consistent with several interpretations as well as the parasite hypothesis. (This problem is general to all good genes models: there are many reasons why females may prefer to mate with males that are larger, more vigorous, and that survive well.) A second kind of observation used to support the parasite hypothesis is the demonstration of heritable variation in parasite resistance. Again, this is necessary but not sufficient: most fitness components show genetic variation but are not increasing evolutionarily because of tradeoffs with other traits. For the parasite mechanism to cause the exaggeration of mating preferences, resistance must not only be heritable but also increasing.

Cross-species comparisons provide an alternative approach to testing the parasite hypothesis. Hamilton and Zuk¹ predicted that more heavily parasitized species would show more extreme male display traits. If such an unexpected correlation appears in spite of the 'noise' produced by unrelated evolutionary processes, it would represent strong evidence for an evolutionary link between parasitism and male displays. Results of the first comparative studies supported the parasite hypothesis. Researchers reported a significant relationship between male coloration and parasite load in species of North American¹ and European⁴³

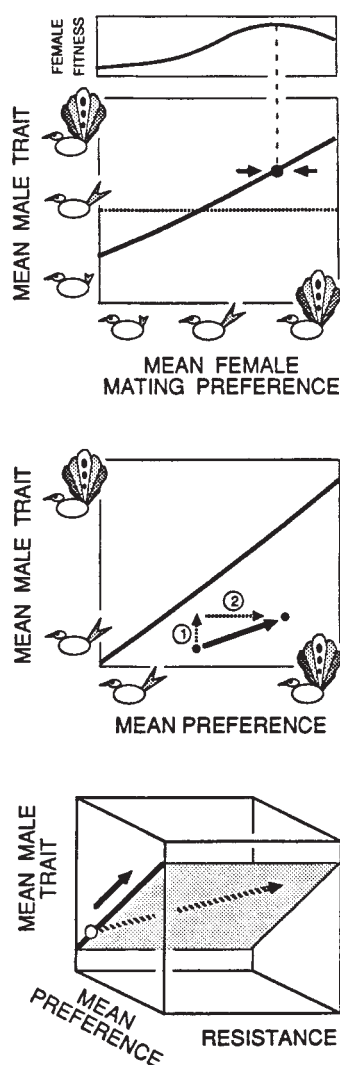


FIG. 4 Three hypotheses for the evolution of preferences. Under direct selection (top), mating preference genes also affect survival or fecundity. This determines the equilibrium for the preference and, through it, the male trait. In a runaway process (middle), the equilibria curve becomes unstable. Selection on the male trait (①) also causes evolution of the female preference through a genetic correlation (②), with the result that trait and preference are exaggerated indefinitely as they evolve away from the equilibria curve. In the parasite hypothesis (bottom), a genetic correlation is established between resistance to parasites and female preferences for a male display trait. Directional evolution of resistance (dashed arrow) results in evolution of the preference, and this causes the preference and trait to coevolve towards greater exaggeration along the equilibria curve (solid arrow).

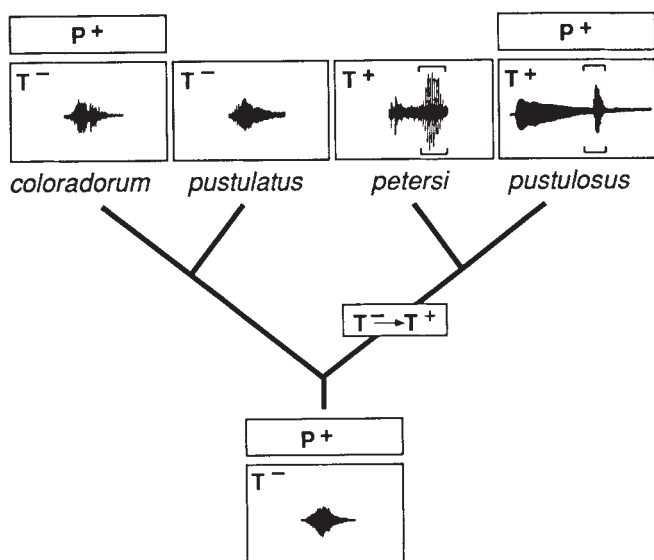


FIG. 5 Male display traits and female mating preferences in *Physalaemus* frogs. Only the closely related *P. pustulosus* and *P. petersi* add 'chucks' to their calls (in square brackets), and so this evolved in their immediate ancestor. Females of both *P. pustulosus* and *P. coloradurum* prefer calls with chucks, from which it is inferred that the common ancestor of all four species (bottom) had the preference. The preference for the chuck therefore evolved before the chuck itself did. (T⁻, no male chuck; T⁺, male chuck; P⁺, female preference for chuck. Oscillogram at bottom is typical of other calls in the genus.)

passerine birds and in British fishes⁴⁴, and also between bird song complexity and parasite load¹. Later studies, however, have suggested that those correlations may be artefacts resulting from the phylogenetic relationships between species and from the methods used to score the male traits. Reanalyses of the North American and European passerine plumages⁴⁵, and bird song complexity⁴⁶, and a new data set on North American fishes⁴⁷ do not find support for the parasite hypothesis.

Conclusions and prospects

Considerable circumstantial evidence has accumulated showing that direct selection is important in the evolution of mating preferences, even in species that form leks and where females receive no material resources from their mates. The data supporting indirect selection of preferences is more equivocal. The runaway hypothesis has the least empirical support, but is also the most difficult to test. Results of most (but not all) within-species studies are consistent with the parasite hypothesis, but are also consistent with alternative interpretations. In the most recent cross-species comparisons, no statistically significant correlations have been found that support the parasite hypothesis. We conclude that the present evidence for direct selection is stronger than that for either the runaway or parasite hypothesis. It is impossible to know whether the lack of strong support for the indirect selection hypotheses stems from the difficulties of testing them or because indirect selection is a weak force in the evolution of mating preferences.

Two major opportunities for empirical studies exist. First, within-species studies need to change focus from the variation in male traits and how it affects male mating success to the variation in female mating preferences. A start has been made by pioneering studies that have identified heritable variation for female mating preferences in several species^{48–50}. Using such systems, one could quantify direct selection and the other evolutionary forces acting on preferences. Second, cross-species studies will continue to give the advantages of a greater taxonomic perspective and longer time scales. The comparative method may be the most fruitful way to study some hypotheses such as the parasite mechanism and to reconstruct the sequence of evolution of traits and preferences.

We close by returning to our opening theme. Recent research has been devoted largely to a few hypotheses for preference evolution. Little attention has been given to the other mechanisms that have been proposed (Table 1), or to how several mechanisms may work simultaneously. Worthy goals for the next generation of field and theoretical studies will be to broaden the search for the forces causing preference evolution, and to study how these mechanisms interact. □

Mark Kirkpatrick and Michael J. Ryan are in the Department of Zoology, University of Texas, Austin, Texas 78712, USA.

- Hamilton, W. D. & Zuk, M. *Science* **218**, 384–387 (1982).
- Bradbury, J. W. & Andersson, M. B. (eds) *Sexual Selection: Testing the Alternatives* (Wiley, Chichester, UK, 1987).
- Kirkpatrick, M. A. *Rev. ecol. Syst.* **18**, 43–70 (1987).
- O'Donald, P. *Genetic Models of Sexual Selection* (Cambridge University Press, 1980).
- Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3721–3725 (1981).
- Kirkpatrick, M. *Evolution* **36**, 1–12 (1982).
- Seger, J. *Evolution* **39**, 1185–1193 (1985).
- Gomulkiewicz, R. S. & Hastings, A. *Evolution* **44**, 757–770 (1990).
- Barton, N. H. & Turelli, M. *Genetics* **127**, 229–255 (1991).
- Fisher, R. A. *The Genetical Theory of Natural Selection* 2nd Edn (Dover, New York, 1958).
- Parker, G. A. in *Mate Choice* (ed. Bateson, P.) 141–166 (Cambridge University Press, 1983).
- Kirkpatrick, M. *Am. Nat.* **125**, 788–810 (1985).
- Andersson, M. *Evolution* **40**, 804–816 (1986).
- Pomiankowski, A. *J. theor. Biol.* **128**, 195–218 (1987).
- Curtis, J. W. & Heisler, I. L. *Am. Nat.* **132**, 437–453 (1988).
- Bulmer, M. *Theor. Pop. Biol.* **35**, 195–206 (1989).
- Heywood, J. S. *Evolution* **43**, 1387–1397 (1989).
- Grafen, A. *J. theor. Biol.* **144**, 473–516 (1990).
- Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
- Zahavi, A. *J. theor. Biol.* **53**, 205–214 (1975).
- Pomiankowski, A. *Oxford Surv. evol. Biol.* **5**, 136–184 (1988).
- Maynard-Smith, J. *J. theor. Biol.* **115**, 1–8 (1985).
- Thornhill, R. *Am. Nat.* **122**, 765–788 (1983).
- Price, T. D. *Evolution* **38**, 327–341 (1984).
- Lightbody, J. P. & Weatherhead, P. J. *Am. Nat.* **132**, 20–33 (1988).
- Ryan, M. J. *Oxford Surv. evol. Biol.* **7**, 156–195 (1990).
- Endler, J. A. in *Speciation and Its Consequences* (eds Otte, D. & Endler, J. A.) 625–648 (Sinauer, Sunderland, Massachusetts, 1989).
- Fleishman, L. *Am. Nat.* (in the press).
- Christy, J. *Ethology* **78**, 113–128 (1988).
- Nevo, E. & Capranica, R. R. *Evol. Biol.* **19**, 147–214 (1985).
- Robertson, J. G. *Anim. Behav.* **39**, 639–645 (1989).
- Markow, T. A., Quaid, M. & Kerr, S. *Nature* **276**, 821–822 (1978).
- Borgia, G. & Collis, K. *Am. Zool.* **30**, 279–285 (1990).
- Møller, A. P. *J. evol. Biol.* **3**, 319–328 (1990).
- Reynolds, J. D. & Gross, M. R. *Am. Nat.* **136**, 230–243 (1990).
- Ryan, M. J. *The Tungara Frog, a Study in Sexual Selection and Communication* (University of Chicago Press, Chicago, 1985).
- Basolo, A. *Science* **230**, 808–810 (1990).
- Ryan, M. J., Fox, J. H., Wilczynski, W. & Rand, A. S. *Nature* **343**, 66–67 (1990).
- Stoner, G. & Breden, F. *Behav. ecol. Sociobiol.* **22**, 285–291 (1988).
- Houde, A. E. & Endler, J. A. *Science* **248**, 1405–1408 (1990).
- Am. Zool.* **30**, 225–352 (1990).
- Boyce, M. S. *Am. Zool.* **30**, 279–285 (1990).
- Read, A. *Nature* **328**, 68–70 (1987).
- Ward, P. I. *Anim. Behav.* **36**, 1210–1215 (1988).
- Read, A. F. & Harvey, P. H. *Nature* **339**, 618–620 (1989).
- Read, A. F. & Weary, D. M. *Behav. ecol. Sociobiol.* **26**, 47–56 (1990).
- Chandler, M. & Cabana, G. *Oikos* (in the press).
- Heisler, I. L. *Genet. Res.* **44**, 133–149 (1984).
- Majerus, M. E. N., O'Donald, P., Kearns, P. W. E. & Ireland, H. *Nature* **321**, 164–167 (1986).
- Sappington, T. W. & Taylor, O. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6132–6135.
- Kirkpatrick, M. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) 41–53 (Wiley, Chichester, UK, 1987).
- Sanderson, N. *Evolution* **43**, 1223–1235 (1989).
- Kirkpatrick, M. *J. theor. Biol.* **119**, 263–271 (1986).
- Charlesworth, B. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) 21–40 (Wiley, Chichester, UK, 1987).
- Rice, W. R. *Evolution* **42**, 817–819 (1988).
- Seger, J. & Trivers, R. L. *Nature* **319**, 771–773 (1986).
- Lande, R. *Evolution* **36**, 213–223 (1982).
- Charlesworth, B. *J. theor. Biol.* **130**, 191–204 (1988).
- Møller, A. P. *Ecology* **71**, 2345–2357 (1990).

ACKNOWLEDGEMENTS. We thank R. Gomulkiewicz, P. Harvey, A. P. Møller, T. Price and M. Zuk for discussions and other workers who provided results before publication.

Quasi-static fault growth and shear fracture energy in granite

D. A. Lockner*, J. D. Byerlee*, V. Kuksenko†, A. Ponomarev‡ & A. Sidorin‡

* US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

† A. F. Ioffe Physical-Technical Institute, USSR Academy of Sciences, Polytechnicheskaya 26, Leningrad, USSR 194021

‡ Institute of Physics of the Earth, USSR Academy of Sciences, Bolshaya Gruzinskaya 10, Moscow D-242, USSR 123810

The failure process in a brittle granite sample can be stabilized by controlling axial stress to maintain a constant rate of acoustic emission. As a result, the post-failure stress curve can be followed quasi-statically, extending to hours the fault growth process which normally would occur violently in a fraction of a second. Using a procedure originally developed to locate earthquakes, acoustic emission arrival-time data are inverted to obtain three-dimensional locations of microseisms. These locations provide a detailed view of fracture nucleation and growth.

EXPERIMENTAL studies of the fracture of brittle rock have been made difficult by the tendency for fault growth to proceed in a violent, uncontrolled manner after fault initiation. Past attempts to stabilize the failure process have concentrated on the development of 'stiff' loading frames which store relatively little elastic energy during the loading cycle¹⁻³. But many rocks store sufficient elastic energy to drive unstable faulting on their own, so fracture stabilization requires the loading system to remove energy dynamically during the failure process⁴. To accomplish this task, we have developed a fast-acting axial control system that adjusts the load applied to the sample to maintain a constant acoustic-emission rate. We are therefore able to study the post-peak-stress region, in which a macroscopic fault forms in the sample, under quasi-static conditions. We have also recorded relative arrival times and amplitudes of acoustic-emission signals occurring in the sample as it is deformed. Inversion of the arrival-time data provides three-dimensional locations of acoustic-emission hypocentres, resulting in detailed imaging of fracture nucleation and propagation. In addition, we have analysed amplitude information to provide estimates of fracture energy and other fault propagation parameters.

Experimental procedure

A right cylinder of intact Westerly granite was precision ground to a length of 190.5 mm and a diameter of 76.2 mm and jacketed in a polyurethane tube. Six piezoelectric transducers (resonance at 0.6 MHz) were attached directly to the rock and used to monitor the high-frequency acoustic emission generated in the sample as it was stressed. Four other transducers were used to measure periodically the P-wave (compressional wave) velocity field during the experiment. The sample was deformed in a triaxial apparatus at a constant confining pressure P_c of 50.0 ± 0.2 MPa. The axial load was applied to the sample with a hydraulic ram controlled by a fast-acting valve and was automatically adjusted to maintain an approximately constant acoustic-emission rate. This required reversing the ram during fracture growth. Axial displacement was measured outside the pressure vessel and circumferential strain was measured with a foil strain gauge attached to the mid-plane of the sample.

Acoustic emission signals were amplified and sent to a six-

channel monitoring system which recorded on digital tape the relative P-wave arrival times at the transducer sites (to $\pm 0.05 \mu\text{s}$), as well as amplitude and first-motion information. The arrival-time data were later inverted to provide three-dimensional locations for the source events, presumably impulsive microcracking events, that emitted the acoustic signals. The recording system and analysis techniques are discussed elsewhere^{5,6}. Arrival-time precision of the monitoring system sets a limit of ± 0.3 mm for the accuracy of determining source locations. In practice, however, other measurement errors further degrade the location accuracies. The primary source of error is a tendency to pick systematically late arrivals for low-amplitude, emergent waveforms⁷. We estimate that location uncertainties for the large-amplitude acoustic-emission events used in this analysis are typically ± 3 mm. This error estimate is supported by the tight clustering of acoustic-emission locations shown in Fig. 2. To aid in the event location analysis, the average P-wave velocity field⁸ was determined independently during the course of the experiment by periodically measuring transit times of artificially generated acoustic pulses travelling along three different ray paths in the sample.

Fault development inferred from AE locations

In Fig. 1, axial stress is plotted as a function of axial displacement. Fault nucleation occurred soon after peak stress, where there was an abrupt reversal in slope. Figure 2 shows locations of acoustic emissions in six sequential time intervals that correspond to the stress intervals indicated in Fig. 1. Events occurring in each time interval in Fig. 2 are shown in two projections. In the upper plot of each pair, the sample is viewed along-strike, so that the $\sim 68^\circ$ dip of the developing fault plane is clearly seen. In the lower plot, the sample has been rotated clockwise by 90° (as viewed from the top) so that the fault plane is seen face-on. Figure 2a shows the relatively diffuse acoustic emission

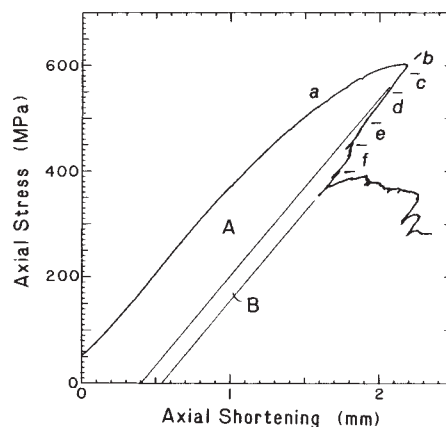


FIG. 1 Axial stress plotted against axial displacement for Westerly granite at $P_c = 50$ MPa. Letters along curve correspond to plots shown in Fig. 2. Region A represents non-recoverable work expended in deforming sample to the point of fault nucleation; region B is work expended during initial fault growth.

that occurs during loading through peak stress. In Fig. 2b, the first 150 s of fault nucleation are plotted, showing that the fault has already established its position and inclined orientation, even though it is only about 20 mm across. The nucleation zone of $\sim 2,000 \text{ mm}^3$ rapidly evolves into the nascent fault, narrowing into a half-penny shape that defines the position and orientation of the fracture. Notice that the fracture nucleated at the sample surface; this was the case for all three granite samples tested so far. In Figs 2c–f, the fracture gradually propagates across the sample as a well-defined band of activity. Relative quiescence in the wake of the advancing fracture front suggests that the band of activity represents a process zone as modelled by Rice^{9,10}. This process zone is 20–30 mm wide in the direction of propagation. Examination in thin section shows that the thickness of the process zone, normal to the fracture surface, is 2–5 mm. Although the fracture grows unimpeded to the upper corner of the sample, it unfortunately propagates downward into the lower steel endplug. As a result, even at the end of the experiment an unbroken column of rock supports part of the axial load. Thus, the sample strength never drops to the frictional sliding strength of $\sim 190 \text{ MPa}$. Although the final stage of fault growth is complicated by the interference of the lower endplug, we infer that nucleation and early growth are little affected. Nucleation occurs near the sample mid-plane, far from the end effects produced by a mismatch in modulus between rock and steel.

Variations in b

The frequency-magnitude relation for earthquakes is commonly described in terms of b , defined by $\log N = a - bM$ where N is number of earthquakes larger than magnitude M and a and b are constants. Analyses of earthquake foreshocks^{11,12} have revealed precursory variations in b . This relation has also been applied in the laboratory to studies of acoustic emission where, for example, b has been calculated for loading of both intact samples¹³ and samples containing faults¹⁴. By assuming that the stress intensity factor at active crack tips would increase as a rock approaches failure, Main and Meredith developed a model

that predicts a drop in b in the pre-failure region followed by a recovery after faulting. They observed this decrease and recovery of b during deformation of porous sandstone. Weeks *et al.*¹⁴ reported similar behaviour for the amplitude of acoustic emission preceding stick-slip on a pre-existing fault in granite.

Because we recorded the amplitudes of the acoustic emission, we can look for variations in b . For each event that triggered the acquisition system, the amplitude of the first peak of the incoming wave train was recorded for each transducer. Amplitudes were then adjusted for $1/r$ geometric spreading and averaged over the six transducers to give an equivalent event amplitude A at a distance of 10 mm from the hypocentre. This is only an estimate of the true event amplitude because it depends on averaging over six directions the effects due to radiation pattern, transducer incidence angle, attenuation and other properties. Referred to input signal level, the acquisition system has a dynamic range of 0.8–17.8 mV which corresponds to a range in A of ~ 5 –120 mV. To analyse variations in b , we have used events occurring in a volume of rock in the shape of a 20-mm-thick disk taken from the central region of the sample. By restricting ourselves to events in this region, we minimize biases due to, for example, attenuation. We assume that $M \propto \log A$, so that b is the slope of the plot of $\log N$ against $\log A$. Representative plots for pre-nucleation, nucleation and propagation events are shown in Fig. 3a. The temporal variation of b is plotted in Fig. 3b. In agreement with Main and Meredith¹³, b decreases to a minimum at the time of fault nucleation. Then, during the propagation phase, b recovers to over 50% of its initial value.

Energy balance and related parameters

In this section we estimate the acoustic energy radiated during fracture propagation. Combining these data with stress and displacement measurements allows us to examine the local energy release rate during fracturing. To calculate acoustic energy we assume that the radiated acoustic-emission energy E_{AE} is proportional to A^2 and also that $E_{AE} \propto E_{in}$ where E_{in} is dissipative, inelastic energy consumed during deformation of

FIG. 2 Sequential plots of locations of acoustic emission. Stress interval for each plot is shown in Fig. 1. Upper plots show events viewed along-strike of eventual fault plane (seen as diagonal band of events). Lower plots show same events when fault plane is viewed face-on. Fault nucleates in b and propagates across sample in c – f . A distinct fracture front develops, in d – f , as fault grows. Number of events per plot in a – f are 474, 123, 402, 1088, 2292 and 4038, respectively.

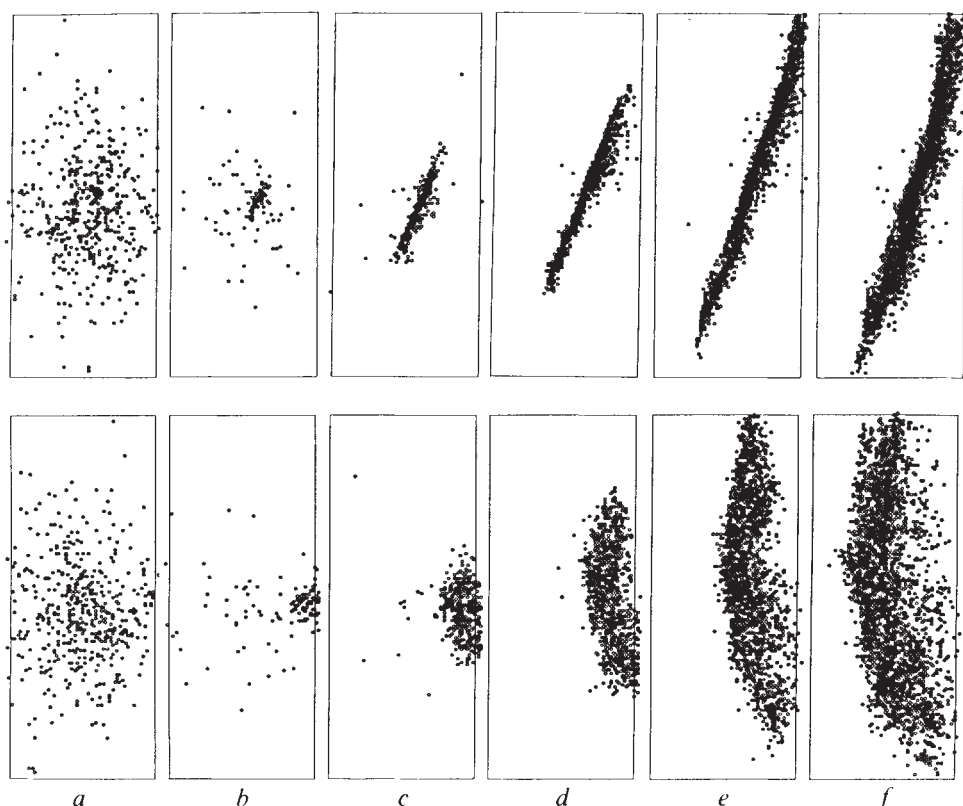
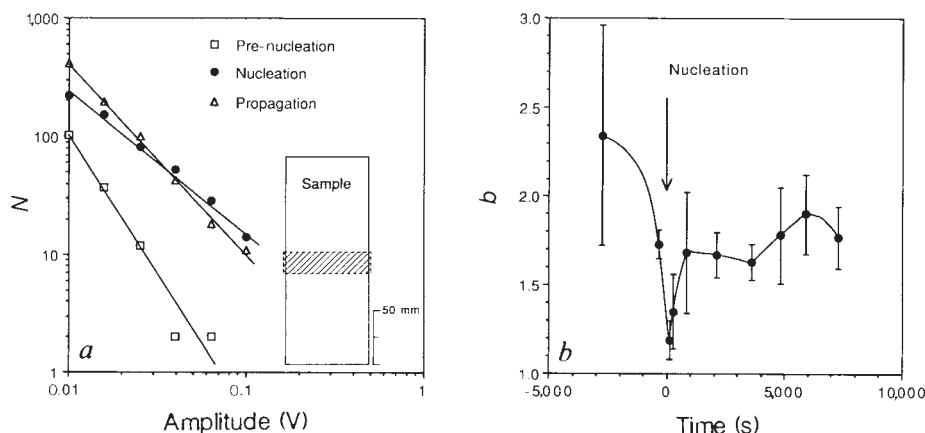


FIG. 3 *a*, Representative amplitude-frequency plots used to determine b . Insert shows region sampled in determining b -value. *b*, Plot of variation of b during experiment with 2σ error bars. Minimum occurs during fault nucleation phase.



the sample. These relations yield $E_{in} = kA^2$ where k is a scaling parameter. Ideally, the transducers would be calibrated so that amplitude would express, for example, displacement. Because we do not know the response function for the transducers, we express their output in volts and include the necessary scaling factor in k . The transducers have a resonance at 0.6 MHz and should be most sensitive to wavelengths of a few millimetres. To the degree that the frequency-amplitude relation is stable, however, the frequency dependence of k will not change. This assumption will be tested in future experiments. In addition to transducer response characteristics, k also includes the acoustic efficiency of the inelastic deformation in the rock. To determine k , we calculate $\int dE_{in}$ and $\sum A^2$ for a portion of the experiment. Inelastic work is calculated in two parts: $\int dE_{in} = \int F_a dl_{in} + \int F_r dr_{in}$. The axial force F_a is $\sigma_a \times$ area of end of sample and the radial force F_r is $P_c \times$ area under jacket; dl_{in} is calculated from the record of axial displacement by removing elastic shortening; dr_{in} is calculated from the record of circumferential strain gauge (not shown). Once the fracture nucleates, deformation becomes inhomogeneous. This will not affect the calculation of axial work because total axial force and displacement are measured. By contrast, the strain gauge provides a local measurement and cannot give the total work expended against the confining medium. The strain-gauge signal is, however, the only information available for the lateral strain. Using the strain-gauge output at face value, the confining pressure provides less than 30% of the total energy flux during fault growth so that the errors due to strain inhomogeneity are not expected to affect the basic results.

Two determinations of k were made using events from pre-nucleation and post-nucleation portions of the experiment (shown as regions A and B in Fig. 1). The post-nucleation determination included fault growth up to the point where the fault ran into the lower endplug. Pre- and post-nucleation calibrations were $3,300 \pm 550$ and $46 \pm 8 \text{ J V}^{-2}$, respectively. The difference of a factor of 70 in the calibrations is larger than the numerical uncertainties in the calculations and indicates a sig-

nificant increase in efficiency of high-frequency acoustic energy generated during fracture propagation. This increase in efficiency following fault initiation is consistent with the observed decrease in b . A decrease in b will result in proportionately fewer low-amplitude events. As the acquisition system does not detect events below the pre-set threshold level (0.8 mV), a decrease in b corresponds to a decrease in the fraction of acoustic energy undetected by the system. Other effects¹⁰, such as a change in the relative amounts of slow, subcritical microcrack growth and rapid, unstable microcracking, may also contribute to the change in calibration.

The lack of detected activity following the passage of the fracture front (Fig. 2*d-f*) may also affect our determination of acoustic efficiency. The decrease of acoustic emission from the newly formed fault surface may be related to the reduction in grain size that occurs as the fracture front passes. These smaller grains should produce acoustic emission with a different frequency content from the unfractured rock and should therefore be detected with a different efficiency by the transducers used. In related experiments we have had no difficulty in detecting acoustic emission from slip on existing faults^{14,15}, suggesting that this effect does not introduce a significant bias. This issue should be clarified in experiments which will be extended to include slip on the fully developed fault.

We can now examine the energy release rate for shear fracturing. Following Rice^{9,10}, we define the breakdown zone at the advancing fracture tip as the region over which shear stress drops from its peak value τ_p to the frictional sliding strength τ_f (Fig. 4). We denote the in-plane width of this zone by w and the total displacement by δ^* . The energy release rate G (Fig. 4) as used in fracture mechanics^{3,9,10}, is the excess, above the frictional energy E_f , expended in the process zone. We first calculate the total energy release rate, denoted by $G_t = E_f + G$, by analysing the radiated acoustic energy and using the post-nucleation calibration of acoustic-emission amplitudes.

In Fig. 5*a* we show time plots of the cumulative acoustic energy A^2 from sources within unit cubes 10 mm on a side. Cube

FIG. 4 Schematic diagram showing slip weakening of rock in process zone. Total energy release rate $G_t = E_f + G$.

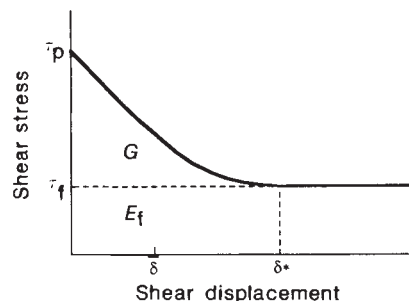
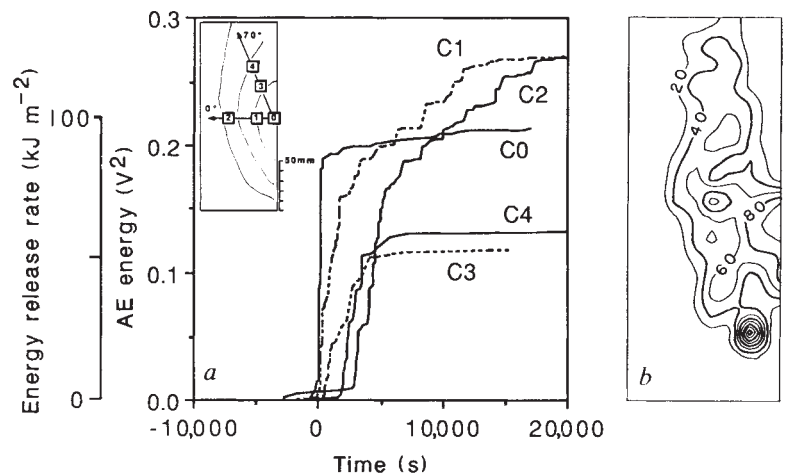


FIG. 5 a, Time plots of cumulative acoustic energy. Calibration of acoustic energy provides estimate of local energy release rate G_t . b, Distribution of G_t on fault surface after 7,000 s of fault growth. Contour intervals are 20 kJ m^{-2} . Variation in energy release indicates local strength heterogeneity.



CO is centred in the nucleation zone. Given the geometry of the fault, cubes C1 and C2 sample pure mode-III anti-plane growth. Cubes C3 and C4 lie along a line inclined 70° to the horizontal. Along this direction the relative contributions of mode-II in-plane and mode-III anti-plane shear are approximately 3 to 1. All samplings show a rapid rise in acoustic energy, marking the arrival of the fracture front, followed by a gradual decrease in the level of acoustic activity. Figure 5a shows a correlation between total radiated acoustic energy and propagation direction. Expressed in terms of total energy release rate, we find $G_t^{\text{III}}/G_t^{\text{II}} \approx 2$. The apparent difference in G_t in the two propagation directions can be checked in the following way. In Fig. 5b we plot the total energy release (kJ m^{-2} per unit area) in the 7,000-s interval following nucleation. At this time, the fracture front has traversed slightly more than halfway across the sample. Although care was taken to choose a uniform, fine-grained (0.2-mm-diameter) sample, the energy release pattern shows distinct local structure which indicates significant strength heterogeneity. Detailed examination of the sequence of acoustic locations shows that fault growth was indeed episodic, as if strong patches periodically retarded local fault growth. The pattern in Fig. 5b shows a generally lower energy release rate in the upper portion of the sample than in the central region. Although this may be related to the direction of fault propagation, it could also be the result of a pre-existing strength heterogeneity. The fault plane did propagate more rapidly into the upper sample region, indicating that this area was relatively easy to fracture. Future experiments should help to clarify this question. Evaluation of the acoustic energy release after 10,000 s gives an average release rate of $G_t \approx 9 \times 10^4 \text{ J m}^{-2}$ with local variation of up to 50%.

Wong³ and Rice⁹ calculated the excess energy release rate G from a plot of shear stress against displacement such as that shown in Fig. 4. In our experiment, however, G_t is probably the more appropriate parameter to consider, because there is no pre-existing fault to which frictional energy E_f can be related.

We will, however, briefly indicate how to modify our results to estimate G . By performing the same calculation as Rice and Wong, we can obtain a rough estimate of G from our experiment. The result is $G = 1.3 \times 10^4 \text{ J m}^{-2}$, in agreement with their range of estimates, $(0.3-7.3) \times 10^4 \text{ J m}^{-2}$. Our estimate of G is based solely on average stress and displacement measurements, but as our sample is larger than the process zone, this calculation is not strictly correct.

To estimate G from our local acoustic energy determinations, we need to determine the local frictional energy E_f expended during fault formation. This requires knowledge of τ_f and δ^* (Fig. 4) occurring on the fault as the process zone moves by. τ_f is known from related experiments and δ^* can be estimated from $^{10} w \approx (2 \text{ to } 3) \mu \delta / (\tau_p - \tau_f)$ where w is the width of the process zone, μ is the shear modulus, and $\delta \approx \delta^*/2$ is the 'characteristic' breakdown slip. Using reasonable values for the parameters results in $E_f = (0.8 \pm 0.3) \times 10^4 \text{ J m}^{-2}$. Because $G = G_t - E_f$, we conclude that within the resolution of this experiment, $G = (8 \pm 5) \times 10^4 \text{ J m}^{-2}$ and $G_t = (9 \pm 5) \times 10^4 \text{ J m}^{-2}$ with a significant portion of the uncertainties related to either local strength heterogeneity or direction of fracture propagation.

By controlling stress to maintain a constant rate of acoustic emission, we have stabilized the failure process in brittle rock. Combined with arrival times and amplitudes of the acoustic-emission signals, this technique can provide a detailed view of fault nucleation and growth. This approach to the study of fault growth opens a number of possibilities for future research in such areas as earthquake dynamics and earthquake prediction. In future experiments we plan to investigate the roles of rock heterogeneity, flaw size and fluid pressure on the nucleation process. This should increase our understanding of the transient phenomena that occur before earthquake rupture. In addition, a more sophisticated treatment of the acoustic-amplitude data should provide new methods for determining the local energetics of fault propagation including energy release associated with strength barriers and propagation direction. □

Received 25 October 1990; accepted 16 January 1991.

1. Wawersik, W. R. & Fairhurst, C. *Int. J. Rock Mech. Mining Sci.* **7**, 561-575 (1970).
2. Wawersik, W. R. & Brace, W. F. *Rock Mech.* **3**, 61-85 (1971).
3. Wong, T. F. *J. geophys. Res.* **87**, 990-1000 (1982).
4. Terrada, M., Yanagidani, T. & Ehara, S. in *Proc. 3rd Conf. on Acoustic Emission/Microseismic Activity in Geol. Structures and Materials* (eds Hardy, H. R. & Leighton, F. W.) 159-171 (Trans-Technical, Clausthal-Zellerfeld, 1984).
5. Lockner, D. A. & Byerlee, J. D. in *Proc. 2nd Conf. on Acoustic Emission/Microseismic Activity in Geol. Structures and Materials* (eds Hardy, H. R. & Leighton, F. W.) 11-25 (Trans-Technical, Clausthal-Zellerfeld, 1980).
6. Lockner, D. A., Byerlee, J. D., Kukensko, V., Ponomarev, A. & Sidorin, A. *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-F.) (Academic, London, in the press).

7. Lockner, D. A. & Byerlee, J. D. *Pure appl. Geophys.* **116**, 765-772 (1978).
8. Lockner, D. A., Walsh, J. B. & Byerlee, J. D. *J. geophys. Res.* **82**, 5374-5378 (1977).
9. Rice, J. R. in *Physics of the Earth's Interior* (eds Dziewonski, A. M. & Boschi, E.) (Italian Physical Society and North-Holland, Amsterdam, 1980).
10. Rice, J. R. in *Proc. 1st Int. Congress on Rockbursts and Seismicity in Mines, Johannesburg 1982* (eds Gay, N. C. & Wainwright, E. H.) 57-62 (Symp. Ser. No. 6, South African Institute of Mining and Metallurgy, Johannesburg, 1984).
11. Von Seggern, D. *Geophys. Res. Lett.* **7**, 637-640 (1980).
12. Jin, A. & Aki, K. *J. geophys. Res.* **91**, 665-673 (1986).
13. Main, I. G. & Meredith, P. G. *Tectonophysics* **167**, 273-283 (1989).
14. Weeks, J. D., Lockner, D. A. & Byerlee, J. D. *Bull. seism. Soc. Am.* **68**, 333-341 (1978).
15. Lockner, D. A. & Byerlee, J. D. *J. geophys. Res.* **82**, 2018-2026 (1977).

Black holes and structure in an oscillating universe

William C. Saslaw*

Astronomy Department, University of Virginia, PO Box 3818,
University Station, Charlottesville, Virginia 22903-0818, USA

IF black holes exist in the contracting phase of a closed universe, they will give rise to a pressure and entropy catastrophe. First, the black holes absorb all the radiation; then their apparent horizons merge, and coalesce with the cosmological apparent horizon. All external observers become internal observers. It is possible that the internal metric of some of the merging black holes will be contracting, and others expanding. I suggest here that the resulting violent inhomogeneities can lead to a re-expansion in a significant portion of the universe. Global re-expansion, prompted by the merging of black holes, may thus begin in a semi-classical rather than fully quantum gravitational era, at densities greater than those at which nucleosynthesis occurs. Surviving black holes and inhomogeneities could initiate the formation of structures such as galaxies in the 'new' universe. The behaviour of such an oscillating universe would differ in detail from cycle to cycle.

Although it is seldom strongly emphasized, there is a problem with models of our Universe that expand forever which is as severe as their much discussed flatness, horizon and galaxy-formation difficulties. This problem may be called the 'transience' problem—we observe our Universe in an atypical state which will never be repeated. An ever expanding universe spends only an infinitesimal portion of its existence in this state containing galaxies and stellar energy sources¹. The weak anthropic principle² attempts to rationalize this result by forcing the Universe to be consistent with our present existence, but it can neither predict the future nor differentiate between open and closed models. Here I consider the alternative supposition that we live in an oscillating universe, and explore the effect of black holes in oscillating models. Whether we live in such an oscillating universe is an observational question that hinges on the value of the density parameter Ω_0 .

Throughout this discussion I will use simple models to illustrate the new points. The radius, R , of the homogeneous isotropic closed Einstein–Friedmann universe with cosmological constant $\Lambda = 0$ evolves according to^{3,4}

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi G}{3} \left(\frac{\rho_{m0} R_0^3}{R} + \frac{\rho_{r0} R_0^4}{R^2} \right) - c^2 \quad (1)$$

where ρ_{m0} and ρ_{r0} are the present matter and radiation (equivalent mass) densities and R_0 is the present radius. Here I will use a present expansion rate $H_0 = \dot{R}/R_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and density parameter $\Omega_0 = \rho_{m0}(8\pi G/3H_0^2) = 2$. This implies a present matter density $\rho_{m0} = 10^{-29} \text{ g cm}^{-3}$, a Hubble time of $20 \times 10^9 \text{ yr}$, and a total mass of the universe $2\pi^2 \rho_{m0} R_0^3 = 6 \times 10^{23} M_\odot$. With $\rho_{r0} = 4.5 \times 10^{-34} \text{ g cm}^{-3}$ from the microwave background at 2.7K, the matter and radiation densities would have been equal at a redshift $z_{eq} = R_0/R_{eq} \approx 2.2 \times 10^4$. Thus from equation (1) in the present matter-dominated era $R_0 = c/H_0 = 1.8 \times 10^{28} \text{ cm}$ and the universe will reach its maximum expansion at $R_{max} = 2R_0$. It will then begin to contract and structures of increasing density will merge and evaporate^{5,6}.

Adding black holes will alter the situation dramatically. These holes could come from the collapse of stars ($\sim 1 M_\odot$), from the evolution of active galactic nuclei ($\geq 10^8 M_\odot$), from collapsing radiation fluctuations, or from relics of a previous cycle of the universe. They need not be very massive. Black holes interact

with the structure of a contracting universe by capturing radiation already present, by emitting gravitational radiation as they merge with one another, and by distorting the global horizons.

As the 'big crunch' approaches, in the radiation-dominated phase, R decreases according to equation (1) as

$$R^2 = 2\beta^{1/2} \tau + c^2 \tau^2 \approx 10^{37} \tau \text{ cm}^2 \quad (2)$$

where $\tau \equiv t_c - t$ is the time remaining to the crunch ($t = 0$ when the cycle begins and $t = t_c$ when it ends) and $\beta = (8\pi G/3)\rho_{r0}R_0^4 = 2.7 \times 10^{73} \text{ cm}^4 \text{ s}^{-2}$ so that $R = 3.2 \times 10^{18} \tau^{1/2} \text{ cm}$.

A black hole increases its mass, M , by absorbing radiation and the small amount of matter dragged in by the radiation as $\dot{M} \approx \rho_{rad} \sigma c$, where $\sigma = 27\pi G^2 M^2 / c^4$ is the radiation capture cross-section for a Schwarzschild black hole. This assumes that black holes accrete independently in a uniform radiation field to estimate the timescale when independent accretion ceases to hold. Noting $\rho \propto R^{-4}$, equation (2) gives

$$M = \frac{2M_0 R_0^2 R^2}{R^2(2R_0^2 + \alpha M_0) - \alpha M_0 R_0^2} \approx \frac{M_0 R^2}{R^2 - \alpha M_0/2} \quad (3)$$

where M_0 is the initial mass and $\alpha \equiv \rho_{r0} R_0^4 \sigma_0 c M_0^{-2} \beta^{-1/2} = (27/2)(3\pi/2)^{1/2} G^{3/2} c^{-3} \rho_{r0}^{1/2} R_0^2 \approx 0.13 \text{ cm}^2 \text{ g}^{-1}$. The mass of an accreting black hole in a contracting universe remains almost constant until $R_A = (\alpha M_0/2)^{1/2} = 2.1 \times 10^{13} r_{0,Sch}^{1/2} \text{ cm}$ and $\tau_A = (81/64)c^{-1} r_{0,Sch}$ s where $r_{0,Sch}$ is the initial Schwarzschild radius (c.g.s. units). Thus $R_A \approx 10^{16} \text{ cm}$, corresponding to $\tau_A \approx 10^{-5} \text{ s}$ and an average radiation density of $\sim 10^{14} \text{ g cm}^{-3}$ for an initial mass of $1 M_\odot$. Then the mass of the black hole increases rapidly. This increase will be modified from the earlier straightforward collisional accretion because the photons do not have time to become redistributed over much more than a Schwarzschild length scale before the crunch occurs.

After contraction to

$$R_s \approx 3N^{1/3} \bar{r}_{Sch} \quad (4)$$

where $\bar{r}_{Sch}$ is the average Schwarzschild radius of the N black holes in the universe and subscript s refers to stable orbits, the last stable photon orbits of the black holes overlap, and most of the photons are absorbed. Energy is transported into the black holes by the contracting global metric. The timescale for local accretion may be only slightly shorter than for global contraction.

The maximum average mass that a single black hole could accrete is the radiation mass of the universe, $M_{univ} = 2\pi^2 R^3 \rho_{rad} \approx 2.6 \times 10^{19} R_0/R M_\odot$, divided by the number, N , of black holes. This gives a maximum average Schwarzschild radius $\bar{r}_{Sch} \approx 1.3 \times 10^{53}/(R_s N) \text{ cm}$ and substituting this into equation (4) gives $R_s \approx 6.2 \times 10^{26} N^{-1/3} \text{ cm}$. For $N(M_0/M_\odot)^{3/2} \leq 2 \times 10^{32}$, $R_s > R_A$ and ordinary accretion will be less important than accretion caused by contraction of the universe. Because $N M_0$ is less than the total baryon mass of the universe, this applies to the current cycle if $M_0 \leq 10^{18} M_\odot$, but it need not apply to other cycles. If $R_s > R_A$, individual black holes boost their mass by the ratio of the radiation density to the initial density in black holes just before merging.

Maintaining conventional homogeneous helium synthesis requires any major change in the global dynamics of the universe to occur at $R_s \leq 10^{19}$, above the temperature of $\sim 3 \times 10^9 \text{ K}$ at which helium and heavy nuclei decompose, implying $N \geq 2.4 \times 10^{23}$. Taking these two limits for R_s and N would give average black holes of $\sim 2 \times 10^5 M_\odot$ at R_s .

Absorption of photons produces a pressure catastrophe. Not only is the radiation density reduced substantially, but the peculiar velocities, V , of the black holes are much diminished. They experience a radiation drag force $-\sigma b T^4 V/c$. Neglecting accelerations of black holes by each other, reasonable until shortly before $R = R_s$, gives $V = V_0(1 - \alpha M_0/2R^2)$ where V_0 is the initial peculiar velocity relative to the background radiation. Therefore the peculiar velocities decrease slowly until R closely

* Also, Institute of Astronomy, Cambridge, UK; and National Radio Astronomy Observatory, Charlottesville, Virginia 22901, USA.

approaches R_s , (or R_A if $R_A > R_s$) and then drop catastrophically. A hole's momentum decreases substantially when it absorbs an amount of radiation (with no net momentum relative to the average background flow) equivalent to about its own mass. Thus both the matter and the radiation pressure drop at R_s .

Rapid pressure loss would at first seem to slow the contraction of the universe according to the Einstein–Friedmann equation (1), after the radiation is converted into the 'conserved' mass of black holes. Moreover, the decreased pressure would satisfy the condition³ that in an oscillating universe with $\Lambda=0$ the pressure as $R \rightarrow 0$ must be less than at $R_{\max}$. Growing inhomogeneity and the dominance of black holes, however, make the situation much more complex. Equation (1) will no longer apply. Time-like geodesics that end in black-hole singularities may satisfy the singularity theorems⁷ of general relativity, which do not necessarily require $R \rightarrow 0$. Merging black holes can release substantial gravitational radiation. This process is just beginning to be understood for the case of the two-body Schwarzschild problem^{7,8} but is less clear in the many-body case where successive mergers could turn much of the mass of black holes back into radiation, temporarily.

A short time later, at $R = R_c$, the apparent horizons of individual black holes coalesce with one another and with the global apparent horizon. The exact time this occurs depends on the 'packing fraction' of the distorted black-hole horizons and may vary for different space-like hypersurfaces. It happens at approximately the time when

$$R = R_c = N^{1/3} \bar{r}_{\text{Sch}} \quad (5)$$

All 'observers' which were outside black holes now find themselves inside merging black holes. This leads to an 'entropy catastrophe' in which they find that the black-hole entropy within their new horizon has suddenly decreased, because this entropy is proportional to the surface area of black holes as seen from infinity. No contradiction with the second law of thermodynamics arises because the observer is in a system for which boundary conditions are changing. This raises interesting new questions about the relations between entropy, structure, constraints on the location of observers, possible Weyl-curvature entropy and possible fundamental time asymmetry in quantum gravitational processes which may reverse the contraction of the metric^{9–11}.

The dynamical behaviour of the universe after this stage becomes even more speculative, but I suggest two arguments that lead to the conjecture that substantial regions of the universe may expand to form the next cycle. First, consider the evolution of a non-rotating black hole from the perspective of a comoving internal observer who sees it collapse to a physical singularity at $r=0$ on a timescale

$$t_{\text{Sch}} = \frac{\pi}{2\sqrt{2}} \frac{GM}{c^3} \approx \frac{r_{\text{Sch}}}{c} \approx 5.5 \times 10^{-6} \frac{M}{M_{\odot}} \text{ s} \quad (6)$$

What happens next? One possibility¹² is that space-time can be extended geometrically through $r=0$ to emerge in another universe which may have a different topology. It is not clear what happens to classical matter or fields in these conditions, or what happens when many singularities coalesce, but it is conceivable that some regions could emerge into a new expanding universe and initiate another cycle.

The second possibility is that quantum mechanical effects cause the internal metric of a black hole to begin expanding away from $r=0$. Because the interior of a black hole behaves similarly to a closed universe, the same quantum processes that have been suggested to cause the universe to bounce^{9–11} should apply within a black hole. To an interior comoving observer, the black hole would seem like an oscillating universe with a period given approximately by equation (6). Combining this with equation (2), a black hole of mass M formed at a radius

$$R \geq 10^{16} (M/M_{\odot})^{1/2} \text{ cm} \quad (7)$$

will have time to undergo at least one or two internal oscillations before $\tau=0$. If all black holes in the universe form with the same mass at the same time, they might oscillate coherently to form a homogeneous Friedmann universe (possibly expanding) when they merge. This seems unlikely, however, because even a small number of black holes oscillating out of phase may produce significant effects during this unstable epoch.

When horizons coalesce at R_c , if some black-hole interiors expand and others contract, the metric becomes violently inhomogeneous over scales of $\bar{r}_{\text{Sch}} (\approx 10^5 M_{\odot}$ in the earlier example). Shrinking regions either turn into new black holes in a generally expanding background or, if gravitational energy can be removed from them before new horizons form, they may become lagging regions of the expansion. As the interiors of different black holes develop independently until they coalesce, one would expect these fluctuations to be essentially gaussian. They would promote galaxy formation.

If detailed calculations become possible, there are some useful constraints on this process. For example, this model requires $R_c \leq 10^{19}$ cm for conventional helium synthesis to occur. For helium production inside expanding black holes, the unknown mechanism that creates a net baryon number must produce about the same entropy per baryon in most of those black holes. This ratio is an unresolved problem, as in the standard Friedmann models. It need not be the same from cycle to cycle. Furthermore, the mass which remains in black holes as the universe starts its expansion cycle must be less than the new net mass in baryons if black holes (the masses of which remain constant in the expansion unless they are small enough to evaporate) do not dominate the subsequent dynamics. In the present cycle of our Universe, for example, this requires that the total mass remaining in black holes M_{bh} be less than $\sim 10^{-5} M_{\text{rad}}$ at $R_c \approx 10^{19}$ cm. This ratio could be a fundamental property of such models. Alternatively it could follow from a weak anthropic principle that only with a sufficiently small residual mass in black holes is the cycle long enough that we occur. This is weaker than the usual weak anthropic principle because cycles with greater M_{bh} have shorter periods and, unless an increased number of such cycles exactly compensates their reduced period, the universe would spend more total time in its long cycles.

Current understanding of the transition from contraction to expansion is so rudimentary that this discussion can only be considered a conjecture. Nevertheless, if many black holes survive in the re-expanding part of the universe, their effects may be observable. Although they do not accrete significantly¹³, resulting inhomogeneity alters the helium and deuterium production at $T \approx 10^9$ K. This also relaxes the constraints on Ω_0 , permitting it to be greater than unity¹⁴.

The main direct effects of black holes may be to promote galaxy formation^{15–21}. If the smoothed average density of matter in black holes is about equal to that of particles, black holes can begin clustering when they dominate the radiation density^{16,21}, at about the time when the red shift is 2×10^4 . Because $\delta\rho/\rho \propto R \propto z^{-1}$ during this period, clusters of about 500 black holes will form from an initial Poisson distribution by the time of matter-radiation decoupling at $z_{\text{dec}} \approx 10^3$. Other initial spatial distributions or a range of black-hole masses could enhance this clustering²². After z_{dec} , particles rapidly accumulate in the clusters which then combine on successively larger scales. This hierarchical gravitational clustering eventually results in galaxies and clusters of galaxies. To produce the observed²³ distribution of galaxies in this manner requires that galaxies be formed by $z \approx 7$ in an $\Omega_0 = 1$ universe²⁴. Similar numerical experiments for an $\Omega = 2$ universe have not yet been done, but clustering is faster for larger Ω_0 . Therefore, for $\Omega_0 = 2$ the observed clustering would probably occur even if galaxies formed at $z \approx 4$ –5, especially as there would be some initial clustering. Thus for clustering to grow proportionally with R and form typical galaxies (with about as much matter in black holes as

in stars) having a mass of $\sim 3 \times 10^{11} M_{\odot}$ by $z \approx 5$, the mass of a typical black hole would have to be $\sim 10^4 M_{\odot}$. (Individual masses could be less if the black holes were formed in clusters.) A galaxy would initially contain about 10^7 of them, but not all of these would remain. Many accumulate in the nucleus, either during merging or by later dynamical friction with stars. Some could be ejected at high velocity by few-body interactions (and remain in the halo or escape) and others could merge to form the supermassive black holes believed responsible for quasars and powerful radio galaxies.

In many models of the Universe, galaxy formation ultimately depends on a rather *ad hoc* hypothesis for an initial linear perturbation spectrum. In the models I have proposed here, there would be a new link between the existence of structure and the large-scale oscillating nature of the Universe. Moreover, in these oscillating universes containing black holes, the formation of structure, as well as the existence of life, always gets another chance. $\square$

Received 4 September 1990; accepted 17 January 1991.

1. Islam, J. N. *The Ultimate Fate of the Universe* 17–51 (Cambridge University Press, 1983).
2. Barrow, J. D. & Tipler, F. J. *The Anthropic Cosmological Principle* (Oxford University Press, 1986).
3. Tolman, R. C. *Relativity, Thermodynamics and Cosmology* (Oxford University Press, 1934).
4. Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* (W. H. Freeman, San Francisco, 1973).
5. Rees, M. J. *Observatory* **89**, 193–198 (1969).
6. Tout, C. A., Eggleton, P. P., Fabian, A. C. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **238**, 427–438 (1989).
7. Hawking, S. W. & Ellis, G. F. R. *The Large Scale Structure of Space-Time* (Cambridge University Press, 1973).
8. Bishop, N. T. *Gen. Relativity Grav.* **20**, 573–581 (1988).
9. Penrose, R. in *Three Hundred Years of Gravitation* (eds Hawking, S. W. & Israel, W.) 17–51 (Cambridge University Press, 1987).
10. Hartle, J. B. & Hawking, S. W. *Phys. Rev. D* **28**, 2960–2975 (1983).
11. Hawking, S. W. in *Three Hundred Years of Gravitation* (eds Hawking, S. W. & Israel, W.) 631–651 (Cambridge University Press, 1987).
12. Lynden-Bell, D. & Katz, J. *Mon. Not. R. astr. Soc.* **247**, 651–661 (1990).
13. Carr, B. J. & Hawking, S. W. *Mon. Not. R. astr. Soc.* **168**, 399–415 (1974).
14. Sale, K. E. & Mathews, G. J. *Astrophys. J.* **309**, L1–L4 (1986).
15. Ryan, M. P. *Astrophys. J.* **177**, L79–L84 (1972).
16. Meszaros, P. *Astr. Astrophys.* **37**, 225–228 (1974).
17. Meszaros, P. *Astr. Astrophys.* **38**, 5–13 (1975).
18. Carr, B. J. *Astr. Astrophys.* **56**, 377–383 (1977).
19. Ipser, J. R. & Price, R. H. *Astrophys. J.* **216**, 578–590 (1977).
20. Carr, B. J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **206**, 315–325 (1984).
21. Carr, B. J. & Rees, M. J. *Mon. Not. R. astr. Soc.* **206**, 801–818 (1984).
22. Itoh, M., Inagaki, S. & Saslaw, W. C. *Astrophys. J.* **356**, 315–331 (1990).
23. Crane, P. & Saslaw, W. C. *Astrophys. J.* **301**, 1–6 (1986).
24. Itoh, M., Inagaki, S. & Saslaw, W. C. *Astrophys. J.* **331**, 45–63 (1988).

ACKNOWLEDGEMENTS. I thank B. Carr for comments and suggestions. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under contract from NSF.

modulation can perhaps be attributed to partly inclined field lines, and the voids may be places where the field becomes perpendicular to the surface.

The remarkable visual observations of Father Secchi in the last century, exemplified by the drawings in his book *Le Soleil*¹, are surprising akin to Fig. 1. Bray and Loughhead² introduced the idea of umbral granulation. Danielson³ adopted the term 'umbral dots' based on the results from Stratoscope I.

These contemporary observations showed umbrae with a sprinkling of relatively bright patches whose size was presumably limited by instrumental resolution and seeing. Beckers and Schröter⁴ quantified umbral dots by measuring their average intensity (~ 0.13 times that of the photosphere in visible light), their diameters (~ 0.16 arcsec without correction for instrument and seeing) and lifetimes ($\sim 1,500$ s). They saw no relation between umbral dots and penumbral filaments. Adjabshirzadeh and Koutchmy⁵ proposed that there is probably continuity between the inner penumbral filaments and umbral dots, but were unable to prove it. Likewise, Soltau⁶ found "faint light-bridges between umbral dots and penumbral filaments". Garcia de la Rose⁷ notes that umbral dots tend to form chains. The results presented here suggest that incipient filamentary structure was being seen in these latter studies.

How can one account for the clarity of filamentary detail seen in Fig. 1? Is there something special about this spot? This seems unlikely, especially as another spot on the same frame shows similar markings (although less distinctly because the seeing

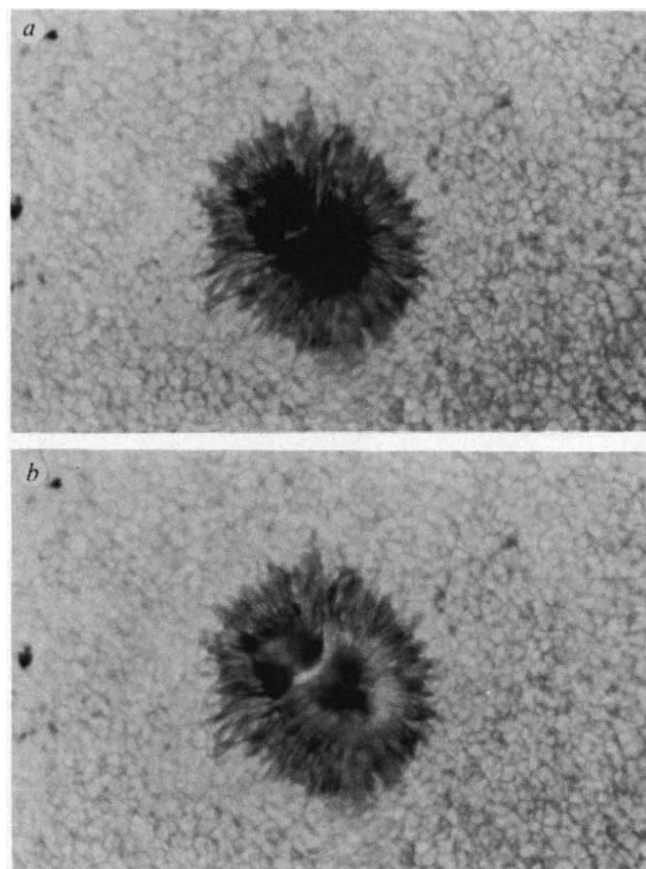


FIG. 1 *a*, Conventional picture of a unipolar sunspot near disk centre, based on a full-disk NSO magnetogram. Suspended over the umbra is a rope-like lightbridge. The outer diameter of the spot is about 32 arcsec. Exposure 1/500 s, as is appropriate for granulation. *b*, Same as *a*, but overprinted with umbral information from an exposure of 1/150 s. Fibrous intrusions delineate still darker areas (voids). Photographic data obtained on 9 September 1990, 1530 UT, West Aux image, aperture 91 cm, focal length 35.8 m, Kodak SO340 20.3 × 25.4-cm film, Wratten 22 filter.

Radial filamentary structure in a sunspot umbra

W. Livingston

National Astronomical Optical Observatories, National Solar Observatory,
PO Box 26732, Tucson, Arizona 85726, USA

THAT the dark umbrae of sunspots are not uniformly dark has long been recognized¹, but our knowledge of umbral structure is still limited. Here I present two white-light photographs of the solar disk, of different exposure, recently obtained during exceptional seeing conditions with the McMath Telescope on Kitt Peak (Fig. 1). The shorter exposure reveals the solar granulation, and the longer brings out details in the darker umbrae. When these two images are suitably combined, the umbra of one small sunspot is resolved into an approximately filamentary structure, connecting to the penumbra, along with featureless voids. A 'light bridge' partially spanning the spot is found to have a broader underlying filamentary foundation that completes the crossing. These photographs suggest that the contemporary view of umbral structure as consisting primarily of granulation or 'umbral dots' is questionable. Given that a sunspot is a place where kilogauss magnetic fields concentrate and diverge outwards, filamentary brightness

was poorer in that region). I believe that filamentary detail was recorded in Fig. 1 simply because there was adequate modulation-transfer-function (MTF) power at the spatial frequencies involved. These umbral features have low contrast. A telescope that resolves high-contrast penumbral structure might still lack the MTF power to provide a good image of an umbra. The resolution in this photograph is then a consequence of the large (91-cm) aperture of the telescope, the fine figure of its optics (B. Graves, personal communication), and the luck of an effective pause of atmospheric turbulence over a part of the image field.

We have combined the penumbra and umbra exposures by tapering the intensity of both images at their junction. I feel that the best published pictures of spot umbrae have been printed in a way that hides the penumbra-umbra connection. Most often the umbral image is either separate or is superimposed as a dodged projection on the penumbral print^{4,5,8}. Alternatively, the pictures have been printed simply for umbral details so that scattered light shrinks the umbral area³. Either way any connecting penumbral structure has been lost. In all the umbral pictures presented by these and other workers⁹⁻¹³, however, the outer umbral boundary is fluted, suggesting that radial filamentary detail was present. Note that ref. 8 contains good examples of the structures that I term voids.

Papathanasoglou¹⁴ has reported observations similar to these: "the bright penumbral filaments penetrate deeply into the umbra and... many cross the umbra joining with filaments on the opposite side". Neither his photographs nor sketches substantiate this idea, however. His observations seem to suffer from seeing-induced reseau patterns, a defect commonly found on small-telescope aperture granulation pictures. Seeing-induced reseau can transform a filamentary image into one of dots—a possible explanation for some of the umbral dot pictures that have been reported. I conclude that umbral dots either do not exist or play only a minor role in umbral structure.

Magnetic measurements within umbrae using the Zeeman-sensitive Fe I 15,648-Å line (D. Rabin and W.L., manuscript in preparation) indicate that field strength increases with a decrease of temperature. This relation means that the voids, which are the darkest and therefore coolest parts, are then the sites (and thus the delineations) of the strongest magnetic fields (the level of which is yet to be determined).

How does the observed filamentary structure change our understanding of sunspots? That knowledge is still rudimentary, with answers still lacking to questions such as why umbrae are so cool (perhaps the result of 'convective inhibition' by magnetic fields?), why there is such a clear intensity discontinuity between the penumbra-umbra boundary, why some spots lack penumbrae, and where the lost radiative energy reappears. Parker¹⁵ proposed a sunspot model in which the umbral dots are field-free areas. These observations now suggest a revised model in which penumbral filaments continue without disruption into the umbra with an associated brightness signature. But more facts are needed before we can achieve a more complete understanding of sunspots.

One reviewer mentioned that Scharmer has made charge-coupled-device (CCD) sunspot movies that have unprecedented resolution (presented at a conference on High Spectral Resolution Solar Observations, Sunspot, New Mexico, 1988), using the new Swedish 50-cm telescope on La Palma. It would be interesting to know if he finds that voids are really featureless or whether they represent just another level in the sunspot atmosphere. □

6. Soltan, D. *Astr. Astrophys.* **107**, 211–213 (1982).
7. García de la Rosa, J. I. in *The Role of Fine-Scale Magnetic Fields on the Structure of the Solar Atmosphere* (eds Schröter, E. H., Vázquez, M. & Wyller, A. A.) (Cambridge University Press, 1987).
8. Grossmann-Doerth, U., Schmidt, W. & Schröter, E. H. *Astr. Astrophys.* **156**, 347–353 (1986).
9. Krat, V. A., Karpinsky, V. N. & Pravidjuk, L. M. *Solar Phys.* **26**, 305–317 (1972).
10. Vázquez, M. *Solar Phys.* **31**, 377–387 (1973).
11. Müller, R. *Solar Phys.* **61**, 297–300 (1979).
12. Moore, R. L. *The Physics of Sunspots* (eds Cram, L. E. & Thomas, J. H.) 259–311 (National Solar Observatory, Sunspot, New Mexico, 1981).
13. García de la Rosa, J. I. *Solar Phys.* **112**, 49–58 (1987).
14. Papathanasoglou, D. *Solar Phys.* **21**, 113–115 (1971).
15. Parker, E. N. *Astrophys. J.* **234**, 333–347 (1979).

ACKNOWLEDGEMENTS. M. Hanna performed the photographic printing. Comments from J. Harvey, R. Howard and N. Sheeley were helpful. NOAO/NSO is operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the NSF.

Experimental and theoretical determination of the magnetic susceptibility of C₆₀ and C₇₀

R. C. Haddon*, L. F. Schneemeyer*, J. V. Waszczak*, S. H. Glarum*, R. Tycko*, G. Dabbagh*, A. R. Kortan*, A. J. Muller*, A. M. Mjlsce*, M. J. Rosseinsky*, S. M. Zahurak*, A. V. Makhija*, F. A. Thiel*, K. Raghavachari*, E. Cockayne† & V. Elser†

* AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

† Department of Physics, Cornell University, Ithaca, New York 14853, USA

THE magnetic susceptibility of C₆₀ and the possibility of magnetic-field-induced π -electron ring currents in this carbon spheroid have been of interest since the initial experiments on carbon clusters¹. If the molecule is regarded as a sphere with a radius of 3.5 Å, on which 60 electrons are free to move, the Pauling ring-current model predicts a ring-current diamagnetic susceptibility 41 times the π -electron ring-current magnetic susceptibility of benzene with the field normal to the plane of the six-membered ring^{2,3}. London theory predicts, however, that the π -electron ring currents in C₆₀ should be weakly paramagnetic or diamagnetic, depending on the relative bond strengths used in the calculation^{2,3}. With the availability of macroscopic quantities of C₆₀ (ref. 4), it is now possible to study experimentally the magnetic properties of the molecule. Here we report on such measurements. We find that the diamagnetism of C₆₀ is small, a result that we attribute to excited-state paramagnetic contributions to the π -electron ring-current magnetic susceptibility. Thus C₆₀ seems to be an aromatic molecule with a vanishingly small π -electron ring-current magnetic susceptibility. We have performed similar measurements on C₇₀, which indicate an appreciable π -electron diamagnetism, consistent with theoretical calculations. We attribute the differences in magnetic properties of these two molecules to their different fractions of five-membered ring structures. The fullerenes may thus constitute a class of compounds of 'ambiguous' aromatic character, traditional measures of which will not provide an adequate classification.

The C₆₀ and C₇₀ samples were obtained by reversed-phase chromatography of fullerite⁴ on octadecylsilanized silica using 40% toluene in isopropanol as eluant. The sample purity was checked by high-pressure liquid chromatography. We measured the magnetic susceptibility of a 0.103-g sample of C₆₀ from 4.2 K to room temperature using the Faraday technique^{5,6}. The measurements were made in an applied field of 1.4 T, and the measured susceptibility was checked for field dependence at several temperatures. The relative accuracy of χ_g is approximately $\pm 1 \times 10^{-8}$ e.m.u.g⁻¹. The absolute accuracy of the susceptibility relative to several standards is 2%.

A paramagnetic upturn was observed at low temperature and was fitted to a Curie-Weiss form, giving a molar Curie constant

Received 5 November 1990; accepted 4 January 1991.

1. Secchi, A. *Le Soleil* (Gauthier-Villars, Paris, 1875).
2. Bray, R. J. & Loughhead, R. E. *Sunspots* (Chapman and Hall, London, 1979).
3. Danielson, R. E. *Astrophys. J.* **139**, 45–57 (1964).
4. Beckers, J. M. & Schröter, E. H. *Solar Phys.* **4**, 303–314 (1968).
5. Adjabshirzadeh, A. & Koutchmy, S. *Astr. Astrophys.* **89**, 88–94 (1980).

TABLE 1 Magnetic susceptibilities of C_{60} and C_{70}

Method	π -Electron ring current (benzene)* (cgs p.p.m.)		Local (cgs p.p.m.)	Total (cgs p.p.m.)	Ref.
C_{60} Expt				-260 ± 20	†
C_{60} London	-0.21	7	-262 ± 40	-255 ± 40	‡§
C_{60} <i>ab initio</i>				-357	¶
C_{70} Expt				-550 ± 50	†
C_{70} London	6.32	-215 ± 50	-306 ± 50	-521 ± 100	†§ #

* Ratio to the π -electron ring-current magnetic susceptibility perpendicular to the plane of the benzene ring, $\chi_M(\pi\text{-benzene-perp})$.

† This work.

‡ Refs 2, 3 and 9.

§ Local (see text and refs 9 and 11). The Hameka method for estimating local magnetic susceptibilities is rigorous for molecules with uniform π -electron charge densities (see ref. 11 and references therein). This condition is fulfilled for C_{60} but not for C_{70} .

|| Calculated with the resonance integrals of all bonds set equal to the benzene value. Note that the calculated value for C_{60} is extremely sensitive to the choice of resonance integrals (refs 2, 3 and 9), and, with the range of likely bond strengths for the molecule, gives 7 to -102 cgs p.p.m. for the ring-current susceptibility in the present parameterization, for a total C_{60} magnetic susceptibility of -255 to -364 cgs p.p.m. It is possible that the magnetic properties of C_{70} will be similarly sensitive to structural features.

¶ Extrapolation of *ab initio* calculation (see ref. 8).

The three components of the π -electron ring-current magnetic susceptibility of C_{70} relative to $\chi_M(\pi\text{-benzene-perp})$ are calculated to be 5.01, 6.98 and 6.98.

$C_M = 0.65 \times 10^{-6}$ e.m.u. per mol C. This small value, corresponding to 1.7×10^{-6} spins per carbon atom, arises from a trace paramagnetic impurity in the C_{60} sample⁷. The final value of the susceptibility was $\chi_M(C_{60}) = -260$ cgs p.p.m. Further measurements were conducted only at room temperature, giving susceptibilities (sample masses) of $\chi_M(C_{60}) = -258$ cgs p.p.m. (0.052 g) and $\chi_M(C_{70}) = -540$ cgs p.p.m. (0.022 g), -565 cgs p.p.m. (0.020 g).

The π -electron ring-current susceptibility was calculated with the finite-field version of the London theory³. For purposes of comparison, all bond lengths were set equal to $1.4 \text{ \AA}$, and idealized geometries based on the structure of C_{60} were adopted. The experimental and theoretical^{2,3,8,9} results are summarized in Table 1, from which it may be seen that there is fairly good agreement between the two for C_{60} .

Given the value of -260 cgs p.p.m. measured for C_{60} , estimates of the π -electron ring-current magnetic susceptibility depend on the value assumed for the local contribution to the diamagnetism^{9,10}. As noted recently¹¹, estimates of these local contributions for the carbon spheroids such as C_{60} are rather uncertain because of the lack of suitable non-planar conjugated model compounds. We adopt the intermediate value of $\chi_M^{\text{local}}(C_{60}) = -262 \pm 40$ cgs p.p.m., which, within the stated error limits, brackets all of the published values^{9,11}. Irrespective of these uncertainties, comparison of the estimated local susceptibility and the measured magnetic susceptibility makes it clear that the π -electron ring current induced in C_{60} must be very small, and probably less than that in a single benzene molecule.

The experimental and theoretical finding of an extremely

small induced ring current in C_{60} poses the question of whether the other carbon spheroids will also be characterized by small π -electron ring-current magnetic susceptibilities. To address this question we have calculated the π -electron ring-current magnetic susceptibility of C_{70} with the finite-field version of the London theory^{2,3} (see Table 1). The London calculation predicts a significant diamagnetism for C_{70} —about 76% of the diamagnetism that would be expected for the 25 benzene rings incorporated in the structure of C_{70} . To compare this calculation with experiment, we require a value for the π -electron ring-current magnetic susceptibility in benzene, together with an estimate of the local contributions. The latter is obtained in the same way as for C_{60} , while for the former we use the value -34 ± 7 cgs p.p.m., based on literature estimates^{9,11,12}. The result is given in Table 1, where it may be seen that the calculated π -electron ring-current magnetic susceptibility of C_{70} amounts to about 41% of the total magnetic susceptibility, compared with the corresponding ring-current contribution (20–25%) to the magnetic susceptibility of benzene^{9,11,12}. The magnetic susceptibility of C_{70} is found to be twice that of C_{60} by both theory and experiment. It should be noted that the calculated π -electron ring-current magnetic susceptibility of C_{70} will probably be subject to rehybridization effects^{2,13} and may be affected by variations in bond length within the spheroid², but without a detailed knowledge of the molecular geometry the importance of these effects cannot be estimated with certainty.

These results suggest that the carbon spheroids will constitute a class of compounds of 'ambiguous' aromatic character¹⁴. That is, the traditional measures of aromatic character are, for these compounds, unlikely to provide a uniform classification of their properties. This is partly a reflection of the non-alternant character¹⁵ of the carbon spheroids¹⁶, but even among alternant hydrocarbons (those without odd-membered rings¹⁷), a unified theory of aromatic character has been formulated only for the Hückel annulenes¹⁸.

To interpret further the ring-current results it is necessary to invoke the usual gauge-dependent separation into diamagnetic (ground-state) and paramagnetic (excited-state) contributions^{2,3}. In the special case of the Hückel annulenes, both components are expressible in terms of the bond order¹⁸. It has previously been argued³, however, that the cancellation of the ring-current diamagnetism in C_{60} arises from the divergence of the Van Vleck paramagnetic term as the relative strengths of the two inequivalent bonds of this molecule are varied. Elser and Haddon³ also drew attention to the presence of the [5]radialene structure in C_{60} , and noted that the ring-current diamagnetism in that five-membered ring is almost completely quenched by paramagnetic contributions. This suggests that the five-membered rings are responsible for quenching the diamagnetism in C_{60} . As the number of five-membered rings is fixed at 12 in the carbon spheroids, with a higher fraction of six-membered rings the larger carbon clusters should show enhanced π -electron ring-current diamagnetism, ultimately approaching the rotationally averaged value for graphite^{9,19}. The results obtained for C_{70} may be interpreted in this light. The stability and electrochemistry²⁰ of C_{60} and C_{70} are similar; our results, however, indicate striking variations in the magnetic properties, and suggest that this should be a rich area for future research.

Note added in proof: Similar results for C_{60} and C_{70} have been obtained by Ruoff *et al.*²¹. □

Received 10 December 1990; accepted 28 January 1991.

- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
- Elser, V. & Haddon, R. C. *Nature* **325**, 792–794 (1987).
- Elser, V. & Haddon, R. C. *Phys. Rev. A* **36**, 4579–4584 (1987).
- Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–357 (1990).
- DiSalvo, F. J., Menth, A., Waszczak, J. V. & Tauc, J. *Phys. Rev. A* **6**, 4574–4581 (1972).
- DiSalvo, F. J. & Waszczak, J. V. *Phys. Rev. B* **23**, 457–461 (1981).
- Tycko, R. *et al. J. Phys. Chem.* **95**, 518–520 (1991).
- Fowler, P. W., Lazzeretti, P. & Zanasi, R. *Chem. Phys. Lett.* **165**, 79–86 (1990).
- Haddon, R. C. & Elser, V. *Chem. Phys. Lett.* **169**, 362–364 (1990).
- Coulson, C. A., Gomes, J. A. N. F. & Mallion, R. B. *Mol. Phys.* **30**, 713–732 (1975).
- Schmalz, T. G. *Chem. Phys. Lett.* **175**, 3–5 (1990).

- Haddon, R. C. *Tetrahedron* **28**, 3635–3655 (1972).
- Haddon, R. C. *Accs chem. Res.* **21**, 243–249 (1988).
- Haddon, R. C., Kaplan, M. L. & Marshall, J. H. *J. Am. chem. Soc.* **100**, 1235–1239 (1978).
- Haddon, R. C. *Aust. J. Chem.* **30**, 1–22 (1977).
- Haddon, R. C., Brus, L. E. & Raghavachari, K. *Chem. Phys. Lett.* **125**, 459–464 (1986).
- Coulson, C. A., O'Leary, B. & Mallion, R. B. *Hückel Theory for Organic Chemists* Ch. 6 (Academic Press, London, 1978).
- Haddon, R. C. *J. Am. chem. Soc.* **101**, 1722–1728 (1979).
- DiSalvo, F. J., Saffran, S. A., Haddon, R. C., Waszczak, J. V. & Fischer, J. E. *Phys. Rev. B* **20**, 4883–4888 (1979).
- Alemand, P.-M. *et al. J. Am. chem. Soc.* **113**, 1050–1051 (1991).
- Ruoff, R. S. *et al. J. Phys. Chem.* (in the press).

Structure of the upper ocean in the western equatorial Pacific

K. J. Richards* & R. T. Pollard†

* Department of Oceanography, The University, Southampton SO9 5NH, UK

† Institute of Oceanographic Sciences, Rennell Centre, Gamma House, Chilworth Research Centre, Chilworth, Southampton SO1 7NS, UK

THE western equatorial Pacific has an important role in the El Niño/Southern Oscillation climate cycle^{1,2}, and sea-surface temperature anomalies in this area are known to affect the global atmospheric circulation³. Little is known, however, about the factors that control the response of the upper ocean, and hence sea surface temperature, to changes in atmospheric conditions. Here we present observations of the density and current structure of the upper layers of the western equatorial Pacific Ocean. We observed frontal features with temperature differences of 1°C extending from the surface down to a depth of 150 m, a strong salinity front associated with a surface convergence, and interleaving of high- and low-salinity waters in 10-m-thick layers that extended for several hundred kilometres in the north-south direction. The large spatial and temporal variability in temperature and salinity observed must influence heat transport, and hence the ocean response to atmospheric forcing (in particular the response to westerly wind bursts, which are thought to be precursors to an El Niño). We estimate the horizontal diffusion coefficient of the interleaving, which indicates that the observed layering may contribute to horizontal mixing within the thermocline.

Until recently it was thought that the western equatorial Pacific was mixed to a depth of ~100 m (ref. 4) and modelling efforts tried to account for this depth⁵. Recent measurements^{6,7} have shown that deep mixing occurs only after a westerly wind burst.

Most of the time the mixed layer is considerably shallower (30 m or less) with a weak but significant stratification below this down to the top of the thermocline at 100–150 m. Surface low-salinity layers up to ~30 m deep have also been observed^{2,6}, which affect the depth to which surface properties can be mixed. Much has been learned about mixing processes in the central Pacific from the Tropic Heat programme⁸, but little is known about what controls the upper ocean in the western equatorial Pacific.

The results we report here are from cruise 32 of the RRS *Charles Darwin* which took place from 3 April to 2 May 1988. The purpose of the cruise was to study the upper density and current structure of the western equatorial Pacific. The principle means of measurement was a combination of SeaSoar (a towed profiling CTD—conductivity, temperature, depth—measurement package) and a hull-mounted acoustic Doppler current profiler (ADCP). Towed at 8 knots, the SeaSoar can measure temperature and salinity with approximately 1-m resolution in the vertical to a depth of 300 m and 4-km resolution in the horizontal.

Two north-south sections, 14 days apart, were made along 165° E from 3° N to 3° S. Each section took two days to complete. The eastward component of the current, the potential temperature and salinity from the sections are shown in Figs 1, 2 and 3, respectively. The period of the cruise was in the 'anti' El Niño phase of the 1986–87 event; a strong undercurrent had been re-established.

Large changes in the current system occurred over the 14-day interval between the sections. The eastward-flowing undercurrent had strengthened and the maximum had moved from 1° S to 1.5° N. In the second section the undercurrent had two cores. (A section taken 14 days later during a cruise sponsored by the United States and the People's Republic of China showed that the undercurrent had a single core centred on the Equator⁹.) The westward-flowing South Equatorial current had also

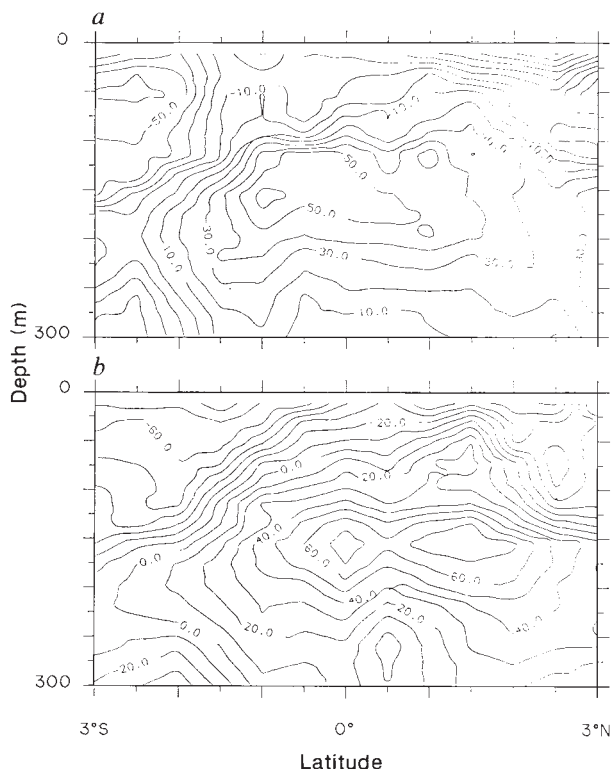


FIG. 1 *a, b*, Eastward component of velocity at 165°E from the two north-south sections taken 14 days apart. The sections are from 3° S to 3° N and from the surface to 300 m depth. The major current features are the westward-flowing South Equatorial current (south of 1° S and between the surface and 100 m depth) and the eastward-flowing Equatorial undercurrent (centred around 150 m depth).

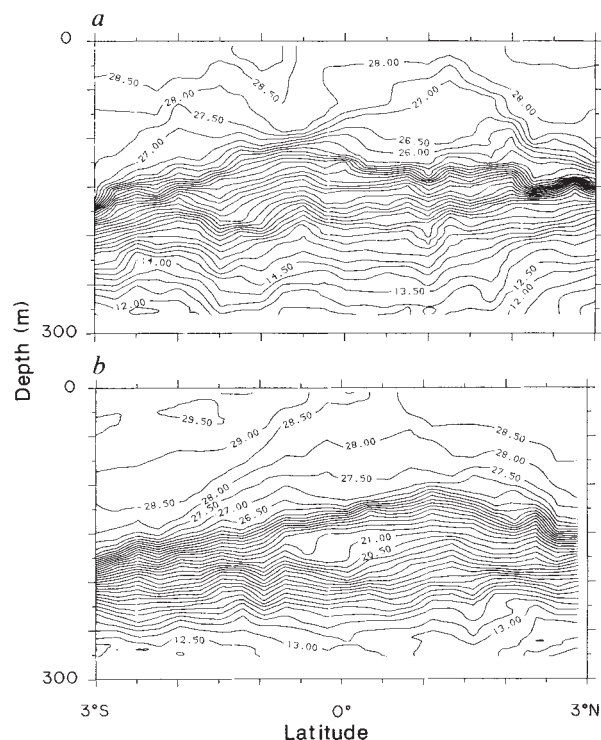


FIG. 2 *a, b*, North-south sections of potential temperature (see Fig. 1 caption). Frontal features with temperature jumps of 1°C are evident in the top 100 m. The main thermocline is between 150 m and 250 m. In *b* there is a large homogeneous lens within the thermocline. Smaller features with reduced temperature gradient are evident in the thermocline of *a*.

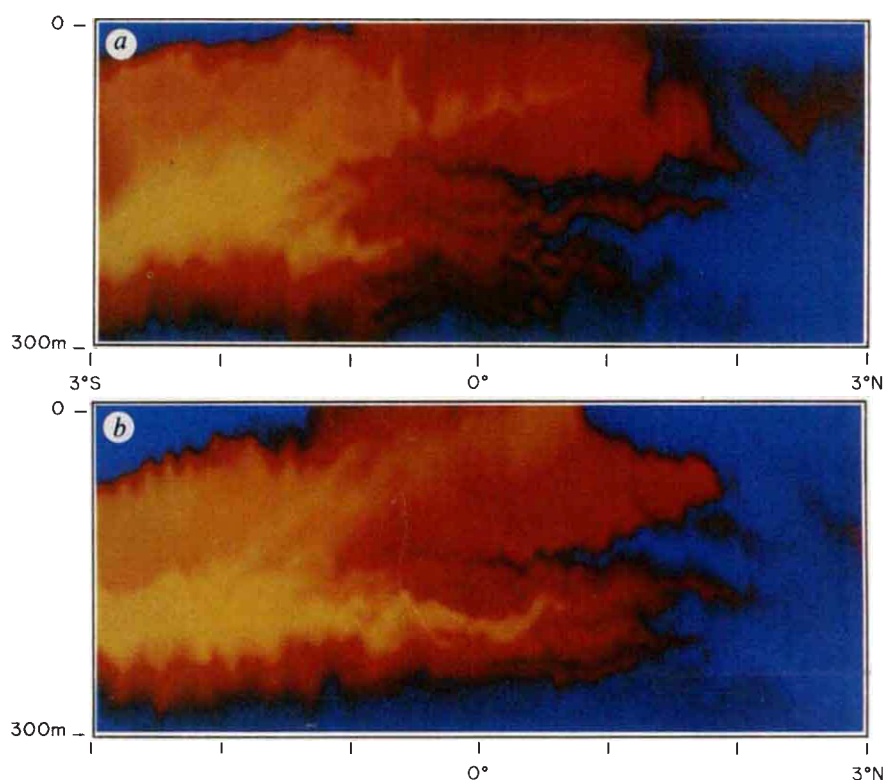


FIG. 3 *a, b*, North-south sections of salinity (see Fig. 1). Blue corresponds to low salinity and yellow to high salinity. The minimum and maximum salinities in p.s.u. (practical salinity units) in *a* and *b* are (34.5, 35.8) and (33.8, 35.9), respectively. The same colour scaling has been used in both images. Interleaving layers extending across the Equator for several hundred kilometres, low-salinity surface layers and a strong surface salinity front are notable features. The amplitude of the salinity variation in the layers can be determined from Fig. 4.

strengthened over the 14-day period and had moved northwards.

The potential temperature had strong gradients in the thermocline below 100–150 m. Above that, the water was still stratified, although more weakly. A number of frontal features were present with temperature changes of 1 °C. The surface temperature south of the Equator had increased by 0.5 °C between the sections. The second temperature section shows a quasi-uniform layer within the undercurrent that was 30 m thick and 100 km across.

The salinity sections show low-salinity surface layers some 30–50 m thick to the north and south of the Equator. The saltier surface waters at the Equator are consistent with upwelling produced by the predominantly easterly winds. Over the period between sections both the northern and southern surface layers had freshened and deepened. The leading edge of the northern layer in the second section had developed a particularly strong front. The distance between successive times when the SeaSoar is at the surface is ~4 km. In this distance the salinity had risen by 0.4 p.s.u. as the ship sailed south out of the low-salinity layer. The position of the salt front coincides with a strong surface convergence that existed over the top 25 m of the ocean. The northward component of the surface current varied from 20 cm s⁻¹, 15 km south of the front, to -20 cm s⁻¹, 15 km north of the front. This large convergence must be compensated for by an equally large divergence in either the east-west or vertical direction. Assuming a vertical divergence distributed over the top 25 m then the vertical velocity would reach 25 m d⁻¹, downwards, in a region where the predominantly easterly winds are expected to produce an upwelling.

The most striking feature in the salinity sections is the interleaving across the Equator of the salty water to the south with the fresher water to the north. These layers were coherent over more than 400 km and were ~10 m thick towards their tips. A number of weaker layers with a vertical wavelength of 30 m are also evident. These seem to be unaffected by the presence of large horizontal shears in velocity as they cross the undercurrent. The scale of the horizontal coherency is much larger than that observed at higher latitudes¹⁰. Vertical profiles of temperature and salinity at the Equator taken from a single dive of the SeaSoar on the first section are shown in Fig. 4. The layers

apparent in the salinity profile are also evident in temperature. The structure shows alternate cooler and warmer layers with reduced and increased salinity, respectively, superimposed on an upwards-increasing temperature gradient. The vertical wavelength of the layers varies from 10 to 25 m.

The Richardson number and stability ratio have been calculated for the two sections. Generally the Richardson number is less than one above the thermocline except for the bottoms of the surface low-salinity layers and the regions of increased temperature gradient. Within the thermocline and undercurrent the flow is generally subcritical, a necessary condition for the interleaving to exist (see Fig. 4).

Over a large part of the thermocline the stability ratio is such as to promote salt fingering. One possible mechanism for the formation of the layers is double diffusive interleaving^{11,12}. The potential energy of the unstable salinity gradient is released by salt fingering to drive the interleaving. The horizontal coherency of the layers in a region of significant horizontal velocity shear suggests that the process is predominantly a north-south interleaving occurring over a large area. Because of the large latitudinal extent of these layers the instability theory for double diffusive instability has been extended to include the equatorial variation of the Coriolis parameter¹³. The increase in the Coriolis parameter away from the Equator is found to limit the latitudinal extent of the layers. Using parameters relevant to the region the theory gives a value of the order of 100 km for the latitudinal extent. The theory also predicts a vertical wavelength of the layers of the order of 10 m. Both values are in reasonable agreement with the observations.

The layering in the thermocline affects not only the vertical mixing through changes in vertical gradients in temperature and salinity but also the cross-Equator fluxes of heat, salt and momentum by the lateral advection of water properties. Assuming a balance between the vertical diffusion of salt and the horizontal advection along a layer¹⁴ we estimate the north-south velocity due to the interleaving to be ~0.02 m s⁻¹. The corresponding effective horizontal diffusion coefficient is $3 \times 10^3 \text{ m}^2 \text{ s}^{-1}$. Such a large value implies that the interleaving is an important horizontal mixing mechanism which may well affect the density and velocity structure of the undercurrent.

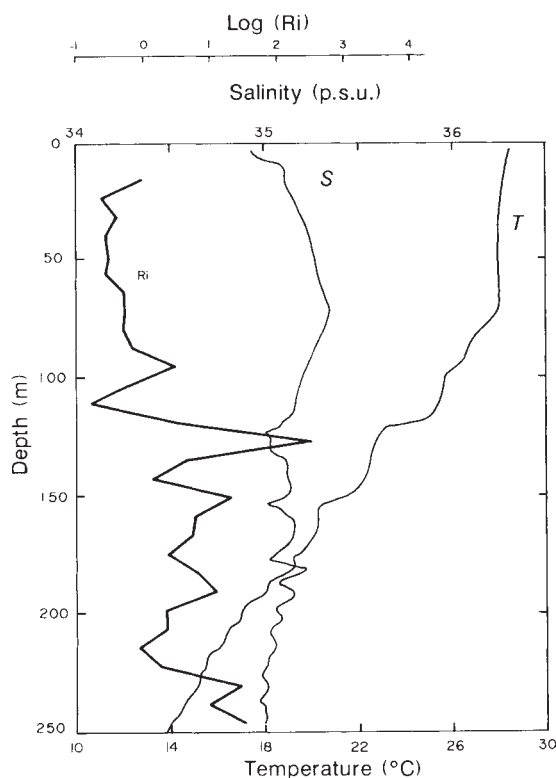


FIG. 4 Vertical sections of temperature T and salinity S taken from a single dive of the SeaSoar at the Equator during the first section. The interleaving below 100 m is evident in both the salinity and temperature. The amplitude of the layers is up to 0.2 p.s.u. in salinity and 1 °C in temperature. Also shown is the gradient Richardson number derived from the density and the ADCP velocity averaged over 8 m in the vertical. Within the thermocline the Richardson number is well over 1.

We have found rich structure in the upper ocean of the western equatorial Pacific, characterized by large spatial and temporal variability in temperature and salinity. The surface low-salinity layer, the 'barrier layer' reported by Lukas and Lindstrom², is very variable in depth and vertical density gradient. We need to understand the processes controlling the density structure of the upper ocean and its variability. Low-salinity layers, fronts, horizontal and vertical advection all play a part, together with interleaving at depth. The observations call into question the ability of present numerical models, which use simple vertical-mixing schemes, to adequately represent the response of the upper ocean to changes in atmospheric conditions—a necessary requirement for the successful prediction of El-Niño events. □

Received 17 August 1990; accepted 7 January 1991.

1. Gill, A. E. & Rasmussen, E. M. *Nature* **306**, 229–234 (1983).
2. Lukas, R. & Lindstrom, E. *Proc. 'Aha Huli' a Hawaiian Winter Workshop on the Dynamics of the Ocean Surface Mixed Layer* (eds Muller, P. & Henderson, D.) 67–94 (Hawaii Institute of Geophysics, University of Hawaii, Honolulu, 1987).
3. Palmer, T. N. & Mansfield, D. A. *Nature* **330**, 533–537 (1984).
4. Colin, C., Henin, C., Hisard, P. & Oudot, C. *Cah. ORSTROM Ser. Oceanogr.* **7**, 167–186 (1971).
5. Garwood, R. W. Jr, Muller, P. & Gallacher, P. C. *J. phys. Oceanogr.* **15**, 1332–1338 (1985).
6. Lindstrom, E. et al. *Nature* **330**, 533–537 (1987).
7. McPhaden, M. J., Freitag, H. P., Hayes, S. P. & Taft, B. A. *J. geophys. Res.* **93**, 10589–10603 (1988).
8. Peters, H. & Gregg, M. C. *Proc. 'Aha Huli' a Hawaiian Winter Workshop on the Dynamics of the Ocean Surface Mixed Layer* (eds Muller, P. & Henderson, D.) 25–45 (Hawaii Institute of Geophysics, University of Hawaii, Honolulu, 1987).
9. Bahr, F., Firing, E. & Songnian, J. *Joint Institute for Marine and Atmospheric Research, Data Rep. no. 5* (University of Hawaii, Honolulu, 1989).
10. Gregg, M. J. *J. phys. Oceanogr.* **10**, 1468–1492 (1980).
11. McDougall, T. J. *J. phys. Oceanogr.* **15**, 1532–1541 (1985).
12. Toole, T. J. & Georgi, D. T. *Progr. Oceanogr.* **10**, 123–145 (1981).
13. Richards, K. J. *J. phys. Oceanogr.* (in the press).
14. Ruddick, B. & Hebert, D. in *Small-Scale Turbulence and Mixing in the Ocean* (eds Nihou, J. & Jamart, B.) 249–262 (Elsevier, Amsterdam, 1988).

Spatial variability in the sink for atmospheric carbon dioxide in the North Atlantic

Andrew J. Watson*, C. Robinson†, J. E. Robinson*, P. J. le B. Williams† & M. J. R. Fasham‡

* Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK

† School of Ocean Sciences, University College of North Wales, Menai Bridge, Gwynedd LL59 5EY, UK

‡ Institute of Ocean Sciences Deacon Laboratory, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK

DIRECT calculation of the air-sea flux of CO_2 requires detailed knowledge of the partial pressure of carbon dioxide (P_{CO_2}) and gas-transfer velocities at the surface of the global ocean¹, with the available observations of surface P_{CO_2} suggesting that it varies in a smooth manner with season and position over the major ocean regions^{2–5}. In spring 1989 we mapped surface P_{CO_2} , total inorganic carbon (TIC), chlorophyll, temperature and salinity at several locations between 47° N and 60° N in the northeast Atlantic near 20° W, observing large variations in P_{CO_2} on spatial scales of <100 km, correlated with plankton chlorophyll, surface temperature and TIC. The variation of P_{CO_2} with latitude was in the opposite sense to that previously reported for this region^{3–5}. Thus, in this ocean area and season at least, the air-sea flux is strongly modulated by biological activity and variable on short spatial scales. The inhomogeneity observed suggests that estimates of the oceanic sink for fossil fuel inferred from existing data (relatively sparse even in the North Atlantic) will be subject to significant error.

Our data, collected as part of the joint global ocean flux (JGOFS) bloom study, consist of high-precision surface P_{CO_2} and TIC measured simultaneously and continuously, together with a comprehensive description of hydrographic and biological conditions from the surface down to a depth of 300 m, obtained by towing a SeaSoar undulating vehicle carrying sensors for temperature, conductivity and fluorescence. Figure 1a shows the ship's track, which consisted of four surveys of ~100 km in extent, one near 47° N, 20° W and three near 59° N, 19° W, with a northward-running transect between these two locations. Atmospheric and surface-water P_{CO_2} were determined by an automated gas chromatograph^{6,7} analysing alternately dried air and gas equilibrated with surface water, whereas TIC was measured by automated coulometric determination of extracted CO_2 ⁸. Surface data were obtained at intervals of ~10 minutes, giving a spatial resolution of ~2.5 km when the ship was under way. Although such complete and continuous carbon system measurements have been undertaken before⁹, the precision of the JGOFS data is considerably higher, as a result of improvements in the measurement technology. The 1σ precisions, obtained from measurements over intervals when the ship was stationary in the water, were $\pm 1.3 \mu\text{atm}$ for surface P_{CO_2} and $\pm 1.3 \mu\text{M kg}^{-1}$ for TIC.

Figure 1b and c shows data taken along the ship's track for the latitudinal transect, and on one of the surveys near 59° N. At all the locations where we surveyed, carbon concentrations at the surface exhibited spatial variability on scales of 10–100 km, correlating with biological activity as determined from the chlorophyll concentration (which was in turn related to the physical structure as revealed by surface temperature changes). The air-sea P_{CO_2} difference ranged from <10 μatm in the coldest waters near 59° N up to 70 μatm in places near 47° N. Horizontal gradients of 5–10 μatm in 10 km were common. The variability and correlation with physical and biological variables that we observed in the open Atlantic is reminiscent of the results obtained by Codispoti *et al.*⁹ in the Bering sea, a high-productive

tivity shelf-sea environment. Contrary to the trend previously reported for the North Atlantic^{3-5,10}, based predominantly on data collected in mid or late summer, we observed the highest air-sea disequilibria towards the south, and a substantial increase in oceanic P_{CO_2} with increasing latitude between 47° N and 60° N. The large air-sea difference towards the south was associated with the occurrence there of blooms dominated by diatoms, whereas to the north there was no sharp spring outburst of productivity during 1989.

Figure 2a and b shows the correlation of P_{CO_2} with surface temperature and chlorophyll concentrations for the entire data set. Like the measurements of Takahashi *et al.*¹⁰ near Iceland, the P_{CO_2} drawdown in these data is inferred to be the result of biological rather than purely physical mechanisms. This is evident both from the locally strong covariation of chlorophyll and carbon, and from the slope of the P_{CO_2} versus temperature

plot (Fig. 2a), which has the opposite sign to that expected if warming of surface water was mainly responsible. Although the evidence is strong that the drawdown is biologically mediated, it is surface temperature rather than plankton chlorophyll that is more consistently related with P_{CO_2} , with the data all fitting on a linear plot of P_{CO_2} against temperature (Fig. 2a). By contrast, although P_{CO_2} usually correlated well with chlorophyll on short spatial scales, no single equation valid for all latitudes related the two. In Fig. 2b, separate regressions have been fitted to the data obtained in the survey near 47° N (which was made in a senescent or secondary bloom), the north-going transect (which passes through the highest chlorophyll concentrations), and the northern surveys where the spring succession was still in its initial stage. Examination of the residuals shows mostly random distribution about the fitted lines except for the transect P_{CO_2} against chlorophyll, which shows the poorest fit and for

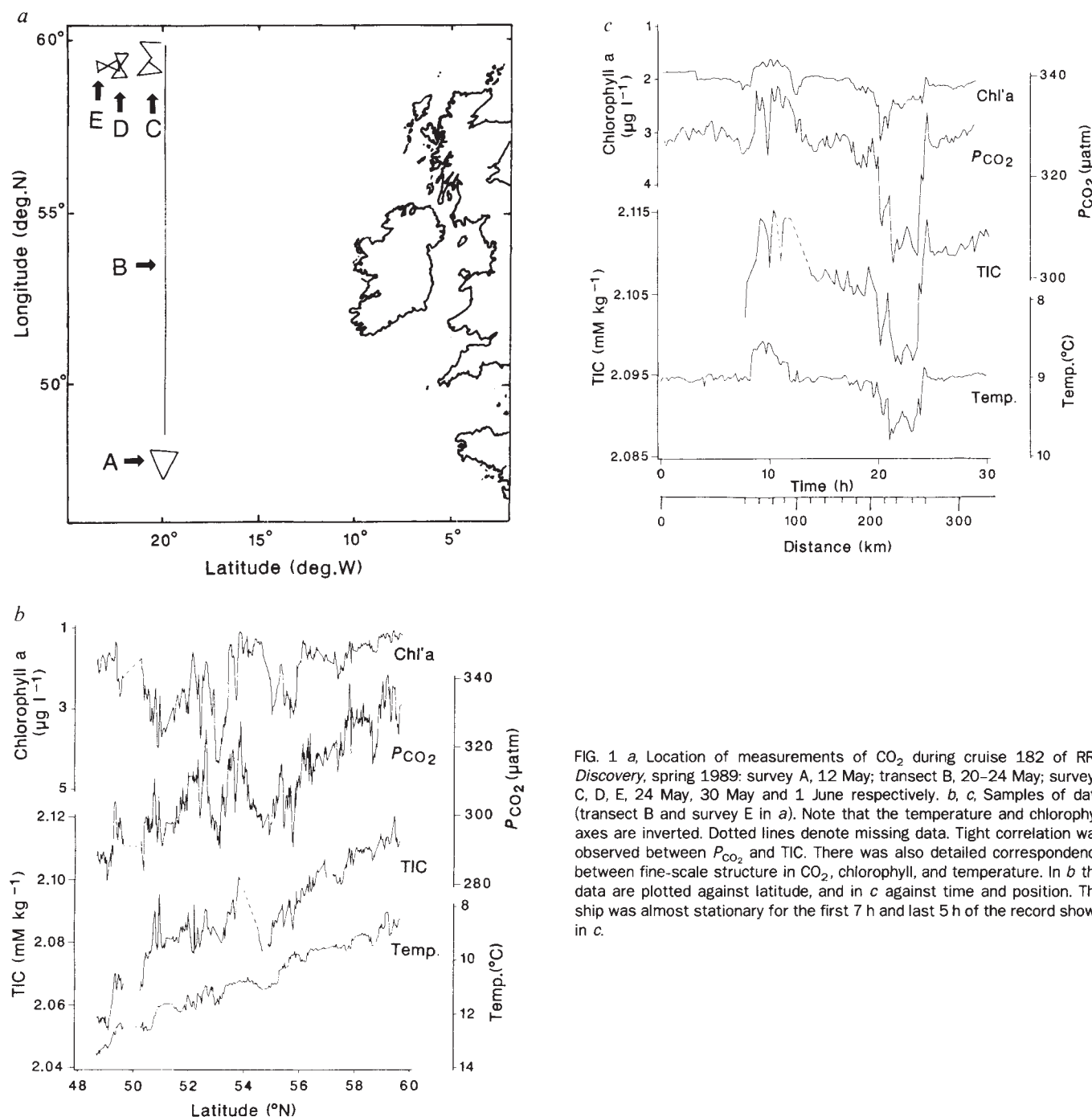


FIG. 1 a, Location of measurements of CO_2 during cruise 182 of RRS Discovery, spring 1989: survey A, 12 May; transect B, 20–24 May; surveys C, D, E, 24 May, 30 May and 1 June respectively. b, c, Samples of data (transect B and survey E in a). Note that the temperature and chlorophyll axes are inverted. Dotted lines denote missing data. Tight correlation was observed between P_{CO_2} and TIC. There was also detailed correspondence between fine-scale structure in CO_2 , chlorophyll, and temperature. In b the data are plotted against latitude, and in c against time and position. The ship was almost stationary for the first 7 h and last 5 h of the record shown in c.

which the northernmost data look similar to the '60° N' line, whereas the southernmost data resemble the '47° N' line.

Our observations are consistent with models¹¹ in which the onset of seasonal stratification, revealed at the surface by an increase in water temperature, is the key event leading to summer drawdown of P_{CO_2} . According to theory¹², stratification of the water initiates net production (and thus net carbon fixation) in the surface layer and also suppresses the upward mixing of deeper water carrying higher carbon concentrations, allowing surface P_{CO_2} values to drop as the phytoplankton abundance increases. After the bloom has depleted the surface of nutrients, the chlorophyll may drop rapidly to quite low values, but P_{CO_2} would be expected to recover more slowly through the summer¹¹ as carbon dioxide invades the surface layer from the atmosphere. We interpret the P_{CO_2} -chlorophyll relation observed near 60° N as being characteristic of a bloom of recent history, whereas the transect and 47° N data are more representative of peak and late stages in the sequence, respectively.

To be consistent with earlier observations¹⁰, the unexpected latitude and temperature dependences we observed must evolve

through the summer, with P_{CO_2} continuing to decrease in the north but increasing in the south, until the latitudinal gradient is reversed and a pattern is established like that previously reported for later in the year. Estimates¹³ and measurements⁵ of the wintertime P_{CO_2} in this region suggest that it lies much closer to equilibrium with the atmosphere, requiring that surface P_{CO_2} must increase substantially during the erosion of the seasonal thermocline in autumn. A simple model of the seasonal progression of a phytoplankton population has been adapted to incorporate carbon dioxide chemistry, and found to reproduce this behaviour well¹⁴.

Recently^{5,15}, attempts have been made to calculate the global net flux of CO_2 into the oceans by combining surface P_{CO_2} data with estimates of gas-transfer velocities across the sea surface. The best estimates of the oceanic sink obtained in this way are lower than required to balance the atmospheric CO_2 budget, by amounts in the range of 0.7–3 Gt C yr⁻¹. In particular, Tans *et al.*⁵ inferred, from an atmospheric model, the existence of a substantial natural sink for CO_2 in the Northern Hemisphere, only a fraction of which was attributable to the ocean sink,

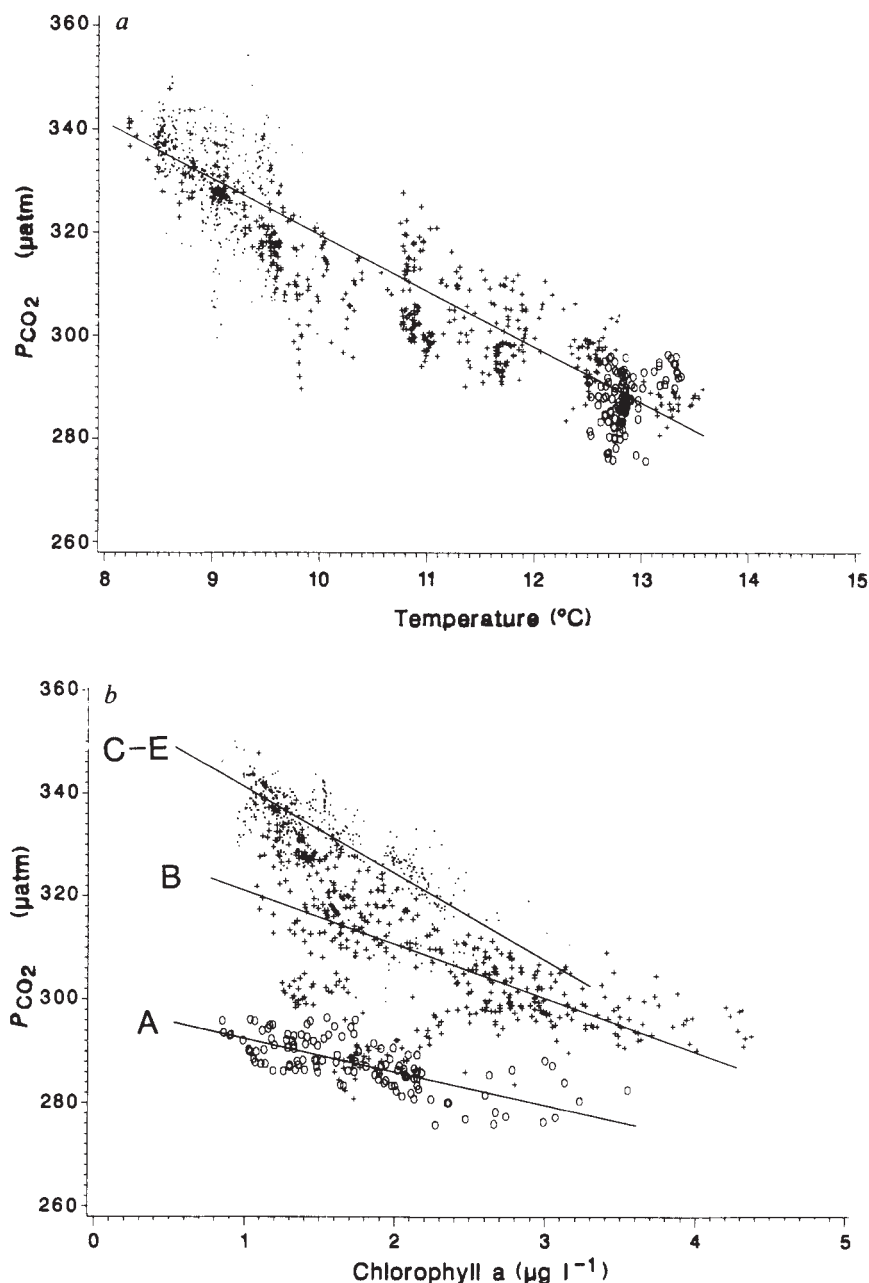


FIG. 2 Plots of P_{CO_2} against surface properties. Data from surveys C-E near 60° N (dots); transect B (crosses) and survey A near 47° N (circles). Note that the average atmospheric concentration determined during the period of the measurements was 355 p.p.m.v. *a*, P_{CO_2} against surface temperature. The regression line (equation; $P_{\text{CO}_2} = 427 - 10.74(\pm 0.16)[T]$, $n = 1,043$, $r = -0.91$, $p < 0.0001$) is fitted to all the data. *b*, P_{CO_2} against surface chlorophyll (Chl). The three regression lines are: for survey A (circles), $P_{\text{CO}_2} = 299 - 6.6(\pm 0.57)[\text{chl}]$, $n = 114$, $r = -0.70$, $p < 0.0001$; for transect B (crosses), $P_{\text{CO}_2} = 332 - 10.6(\pm 0.64)[\text{chl}]$, $n = 475$, $r = -0.61$, $p < 0.0001$; for surveys C-E (dots), $P_{\text{CO}_2} = 358 - 16.8(\pm 0.60)[\text{chl}]$, $n = 450$, $r = -0.79$, $p < 0.0001$.

calculated on the basis of a compilation of spot measurements of P_{CO_2} since 1972. They suggested that the difference must be due to net uptake of CO_2 occurring on land surfaces. Although these calculations of the ocean flux are based on the best available data and the North Atlantic is one of the better sampled oceans, our observations suggest that even there it may be difficult to estimate adequately regional mean P_{CO_2} . We note that, globally, a $1 \mu atm$ error in the mean annual difference between air and sea P_{CO_2} induces an uncertainty of $\sim 0.2 Gt C yr^{-1}$ in the calculated ocean sink, whereas in our data, the 1σ variation at any given point, due to natural variability, is $\sim \pm 10 \mu atm$. This suggests that several tens of such measurements would need to be averaged in order to provide each independent point in a global spatial and temporal survey adequate to define the flux to a small fraction of a gigaton. About 30–40% of the global ocean (and a greater fraction of global productivity) lie in temperature or polar regions subject to an annual bloom, where variability similar to that observed in the North Atlantic might be expected. The problem is compounded by the presence of interannual variations¹⁶ in the timing and intensity of biological activity, in the North Atlantic at least. Hence, although the total (terrestrial-plus-ocean) sink for excess CO_2 can be calculated with confidence from mass-balance considerations⁵, we believe that estimates of how this is apportioned between land and sea are very uncertain at present.

One approach to refining these calculations would be to use satellite-derived observations to make regional and seasonal predictions of P_{CO_2} , using relationships like those shown in Fig. 2 to relate remotely sensed chlorophyll and temperature fields to surface carbon. (In practice, given the ephemeral nature of

the P_{CO_2} -chlorophyll relationship, it may be necessary to use chlorophyll data to constrain regional calculations of the biological drawdown rather than to calculate P_{CO_2} directly.) To be global in application, this approach would require 'sea truth' data from many ocean regions to calibrate the remote-sensing images. We are presently conducting an investigation to derive 'remotely sensed P_{CO_2} ', using the data reported here to calibrate satellite data from the springtime northeast Atlantic, the first results of which are encouraging. □

Received 21 June 1990; accepted 21 January 1991.

1. Thomas, F., Perigaud, C., Merlivat, L. & Minster, J.-F. *Trans. R. Soc.* **A325**, 71–83 (1988).
2. Keeling, C. D. *J. geophys. Res.* **73**, 4543–4553 (1968).
3. Broecker, W. S. *et al. J. geophys. Res.* **91**, 10517–10527 (1986).
4. Takahashi, T. *Oceanus* **32**, 22–29 (1989).
5. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).
6. Takahashi, T., Chipman, D., Schechtman, N., Goddard, J. & Wanninkhof, R. *Lamont-Doherty tech. Rep. Measurements of the partial pressure of CO_2 in discrete water samples during the North Atlantic Transient tracers in the Oceans project* (Columbia University, New York, 1982).
7. Weiss, R. F. *J. chromatogr. Sci.* **19**, 611–616 (1981).
8. Johnson, K. M., Sieburth, J. McN., Williams, P. J. le B. & Brandstrom, L. *Mar. Chem.* **21**, 117–133 (1987).
9. Codispoti, L. A., Friederich, G. E., Iverson, R. L. & Hood, D. W. *Nature* **296**, 242–245 (1982).
10. Takahashi, T. *et al. J. Mar. Res. Inst. Iceland* **9**, 20–36 (1985).
11. Peng, T.-H., Takahashi, T., Broecker, W. S. & Olafsson, J. *Tellus* **39B**, 439–458 (1987).
12. Sverdrup, H. J. *Cons. perm. Int. Explor. Mer.* **18**, 287–295 (1953).
13. Brewer, P. G., Goyet, C. & Dyrssen, D. *Science* **246**, 477–479 (1989).
14. Taylor, A. H., Watson, A. J., Ainsworth, M., Robertson, J. E. & Turner, D. R., *Global biogeochem. Cycles* (in the press).
15. Etcheto, J. & Merlivat, L. *J. geophys. Res.* **93**, 15, 669–15,678 (1988).
16. Williams, R. *Hydrobiologia* **167/168**, 151–159 (1988).

ACKNOWLEDGEMENTS. We thank R. Ling, D. R. Turner, D. Chipman and J. Goddard for advice and assistance with measurements, and officers and crew of RRS *Discovery*, cruise 182 for assistance at sea. This research forms part of the Biogeochemical Ocean Fluxes Study (BOFS), which is the UK contribution to JGOFS. It was funded by NERC, with contributions from National Power.

Importance of continental margins in the marine biogeochemical cycling of carbon and nitrogen

John J. Walsh

Department of Marine Science, University of South Florida,
140 Seventh Avenue South, St Petersburg, Florida 33701, USA

THE continental margins occupy less than 20% of the surface area of the world ocean, and it is widely assumed that they do not play a significant part in the oceanic biogeochemical cycles of carbon and nitrogen. Data from 32 sediment-trap moorings, 16 in the deep sea and 16 on the continental slope¹, suggest that at an average depth of 2,650 m on the slope, the combined rain of surviving shelf and slope particles yields a mean carbon flux of $6.9 g C m^{-2} yr^{-1}$ —about ten times that at the same average depth in the deep sea ($0.8 g C m^{-2} yr^{-1}$). Because the area of the deep sea is about ten times greater than that of the continental slopes, using the sediment-trap data and assuming a carbon/nitrogen ratio of 5:1, the equivalent total particulate offshore nitrogen loss is $0.5 \times 10^{14} g N yr^{-1}$ at 2,650 m. If these trap observations are generally representative of the oceans and continental margins, then the supply of dissolved nitrate to the overlying euphotic zones should also be similar. Here I provide an independent estimate of the annual supply of onwelling nitrate from the deep sea to the shelves and find that it may balance the offshore flux of carbon, suggesting that the continental margins and deep sea are equally important in the carbon and nitrogen biogeochemical cycles.

The mass flux of particulate matter caught with sediment traps increases with depth on the continental slope^{2–4}, implying a lateral transport from the adjacent shelf. If losses of dissolved organic nitrogen back to the slope, the burial of organic matter within the shelf sediments, and evasion of gaseous nitrogen (N_2 , N_2O) to the atmosphere are all small, then the export flux of particulate nitrogen at the shelf break can be estimated from

the exogenous dissolved nitrogen input and the phytoplankton uptake, and is equivalent to 'new' production⁵. Here I ignore fixation of atmospheric N_2 by cyanobacteria and rainfall, as well as freshwater supplies of organic matter. I have also ignored shelf nitrification because it is an endogenous source of nitrate, similar to recycling of ammonium, urea and amino acids within the shelf ecosystem. First I estimate the onwelling flux of dissolved nitrate to continental shelves from the deep sea.

A shelf area of $2.6 \times 10^7 km^2$ and an average width of 85 km yield a mean length of $\sim 3.1 \times 10^5 km$. In an attempt to account for the spatial inhomogeneity of nutrient exchange, I have partitioned shelf regions into those of coastal upwelling in eastern boundary currents, those of cyclonic eddy upwelling in western boundary currents, and those of estuarine-induced exchange (Table 1). The effect of both coastal and eddy-induced upwelling is to transport nitrate from deeper levels of slope waters to shallower depths, for example, to the 100-m isobath at the shelf break. Knowledge of the nitrate stocks at a depth of 100 m along the shelf break and the lateral flows onto the outer shelf yields the advective onshore flux of nitrate to these types of shelves (Table 2). This may be an underestimate, because I have ignored horizontal mixing in these calculations.

When coastal upwelling prevails in eastern boundary currents, the nitrate contents are $30 \mu mol NO_3 l^{-1}$ at 100 m off Oregon⁶, $25 \mu mol NO_3 l^{-1}$ at this depth off Baja California⁶ and southwest Africa⁷, $25–30 \mu mol NO_3 l^{-1}$ off Peru⁸, $>20 \mu mol NO_3 l^{-1}$ off Somalia⁹, and $15 \mu mol NO_3 l^{-1}$ off northwest Africa¹⁰. During 2–3 months of upwelling off Oregon and Peru, the onshore flow past the 100-m isobath is $1–5 cm s^{-1}$ over the lower half of the water column¹¹. The thickness of the benthic boundary layer in these two shelves is calculated to be 10–13 m, when stratification of the water column is considered¹¹.

Coastal upwelling is, however, seasonal; for example, the nitrate content at the 60-m isobath off Oregon is $>20 \mu mol NO_3 l^{-1}$ during only six months of the year¹². Over 12 months, the mean onshore flow within the benthic boundary layer at the 100-m isobath of coastal upwelling regions might be only

TABLE 1 Partitioning of shelf-break length into regions

Shelf-break length (km)	Coastal upwelling	Eddy upwelling	Estuarine exchange*
Northern Hemisphere (0–60°)	36,400	30,000	65,200
Arctic (>60°)	—	—	46,400
Southern Hemisphere (0–60°)	50,000	35,600	14,800
Antarctic (>60°)	—	—	36,000
Total	86,400	65,600	162,400

* The regions of estuarine dispersion are determined by the extent of the 33.0-p.s.u. isopleth of surface salinity, which is similar to ~2.0-p.s.u. deviations from average latitudinal salinities³⁸.

~2.5 cm s⁻¹ with a mean nitrate stock of 20 $\mu\text{mol NO}_3 \text{ l}^{-1}$. Over the coastal upwelling region of $8.6 \times 10^7 \text{ m}$ length (Table 1), a daily onshore flux of $2.5 \times 10^3 \text{ m day}^{-1}$ (20 mmol $\text{NO}_3 \text{ m}^{-3}$) within a 10-m depth interval yields an annual nitrogen input of $2.2 \times 10^{14} \text{ g N yr}^{-1}$ (Table 2).

When cyclonic eddies are present in western boundary currents, the nitrate content at 100 m is 15–20 $\mu\text{mol NO}_3 \text{ l}^{-1}$ in the Florida¹³, Kuroshio¹⁴ and Loop¹⁵ currents. At other times, the content at this depth is at least 10 $\mu\text{mol NO}_3 \text{ l}^{-1}$ in the Brazil current¹⁶, suggesting a mean content of ~15 $\mu\text{mol NO}_3 \text{ l}^{-1}$ (Table 2). Eddy-induced upwelling¹³ is a sporadic process, however, which introduces nutrients to the shelf break only in the lower 20–30 m of the water column. Mean onshore flows of 0.5–1.1 cm s⁻¹ have been observed at 3 m above the 40-m and 75-m isobaths over a four-month period in the South Atlantic Bight¹⁷. At mid-depth on the 75-m isobath, mean onshore flows of 3.0 cm s⁻¹ have also been observed here. Assuming again an onshore flow of 2.5 cm s⁻¹ in a 10-m-thick layer with a nitrate content of 15 $\mu\text{mol NO}_3 \text{ l}^{-1}$ yields an onwelling flux of $1.3 \times 10^{14} \text{ g N yr}^{-1}$ for eddy upwelling (Table 2).

In contrast to the upwelling ecosystems, the nitrate content at 100 m off estuarine-influenced shelves in the North Sea¹⁸, the South China Sea¹⁹, the Beaufort Sea²⁰ and the Mid-Atlantic Bight²¹ is 10 $\mu\text{mol NO}_3 \text{ l}^{-1}$. The diabathic circulation of these regions is more complex, with onshore flows at intermediate depth, rather than in a benthic boundary layer. Long-term moorings at the shelf break on the 198-m isobath south of Martha's Vineyard²² show mean onshore flows of 1.0–2.3 cm s⁻¹ at depths of 88–118 m. With a boundary condition of 10 $\mu\text{mol NO}_3 \text{ l}^{-1}$, the shoreward influx of nitrate to estuarine-influenced shelves at 2.5 cm s⁻¹ over a 10-m-thick layer is estimated to be $2.1 \times 10^{14} \text{ g N yr}^{-1}$ (Table 2).

The total input of deep-sea nitrate across the shelf break is $5.6 \times 10^{14} \text{ g N yr}^{-1}$, compared to a terrestrial flux of perhaps

TABLE 2 Nitrate flux onto continental shelves from the deep sea

	Coastal upwelling	Eddy upwelling	Estuarine exchange
Nitrate content, N , at 100 m ($\mu\text{mol NO}_3 \text{ l}^{-1}$)	20	15	10
Advective flux, uN^* ($10^8 \text{ g N m}^{-2} \text{ yr}^{-1}$)	2.5	1.9	1.3
Shelf-break exchange area, A (10^8 m^2)	8.6	6.6	16.2
Nitrate input, uNA ($10^{14} \text{ g N yr}^{-1}$)	2.2	1.3	2.1

* where u is a mean onshore flow of 2.5 cm s⁻¹.

$0.6 \times 10^{14} \text{ g N yr}^{-1}$ across the shoreline (Fig. 1) from dissolved inorganic nitrogen (DIN) in present river runoff and sewage discharge¹². An estimated terrestrial dissolved-organic-carbon (DOC) loading²³ of $0.8 \times 10^{15} \text{ g C yr}^{-1}$ and a higher-plant carbon/nitrogen ratio of at least 16:1 suggest a similar freshwater input of dissolved organic nitrogen (DON) of $0.5 \times 10^{14} \text{ g N yr}^{-1}$. The fate of terrestrial DON is presumably oxidation on or near the continental margins, but the lability of these compounds is unknown and I have therefore ignored them.

Before the Industrial Revolution, inputs of DIN from the land may have been ten times less than they are today²⁴. Similarly, anthropogenic inputs of atmospheric nitrogen may now be affecting the nitrogen/phosphorus ratios of offshore waters²⁵, but rainfall is still considered a small source of nitrogen to either the coastal zone²¹ or the deep sea²⁶. I have also ignored burial of organic nitrogen on the shelf in this budget. Finally, previous estimates of denitrification^{27,28}, with evasion mainly of N_2 (perhaps 8% is N_2O) from the shelves, are similar at a maximum of $0.5 \times 10^{14} \text{ g N yr}^{-1}$ (Fig. 1), although recent measurements raise the possibility of a tenfold larger flux²⁹.

The present anthropogenic DIN loading of $0.6 \times 10^{14} \text{ g N yr}^{-1}$ may offset a denitrification loss of $0.5 \times 10^{14} \text{ g N yr}^{-1}$, so that $5.7 \times 10^{14} \text{ g N yr}^{-1}$ may be available for 'new' production and export (Fig. 1) if the onwelled nitrate is efficiently converted to particulate or dissolved organic nitrogen. With a mean shelf depth of 50 m, the average residence time of onwelled nitrate would be 167 days, over which light limitation could be significant on shelves at mid-latitudes. With a phytoplankton carbon/nitrogen ratio of 5:1, this dissolved nitrogen import thus allows a maximum carbon export of $2.9 \times 10^{15} \text{ g C yr}^{-1}$ from the shelves in the form of particles and dissolved organic matter.

A mean DOC excretion rate¹² by shelf phytoplankton of 17% of the total photosynthesis of $5.2 \times 10^{15} \text{ g C yr}^{-1}$ suggests that $0.9 \times 10^{15} \text{ g C yr}^{-1}$, or $1.8 \times 10^{14} \text{ g N yr}^{-1}$, might be exported in dissolved form. Perhaps a third of this marine DON loss is replaced by the terrestrial DON import, but the residence time of the two nitrogen pools may differ³⁰. Some unused deep-sea nitrate may exit as well, depending on the season and depth of exodus; after phytoplankton blooms, however, shelf surface waters¹² usually contain less than 1 $\mu\text{mol NO}_3 \text{ l}^{-1}$.

A total particle rain of $2.2 \times 10^{14} \text{ g C yr}^{-1}$ trapped at 2,650 m (spread over a slope area of $3.2 \times 10^7 \text{ km}^2$) and a particulate export of $2.0 \times 10^{15} \text{ g C yr}^{-1}$ at the 100-m shelf break imply that perhaps 10% of the detritus escapes oxidation to dissolved CO_2 during the 2,550-m descent in slope waters. Within 300 km of the shelf break, however, bottom metabolism of margin sediments³¹, where half of the ocean's degradable organic carbon is stored³², suggests that sediment traps may underestimate the influx of detritus by 200–300%, missing the lateral contribution of near-bottom suspended matter. As much as 20–30% of shelf export may thus be stored, or degraded, in slope sediments.

A total primary production of $200 \text{ g C m}^{-2} \text{ yr}^{-1}$ over all shelves (requiring $10.4 \times 10^{14} \text{ g N yr}^{-1}$) and full utilization of the onwel-

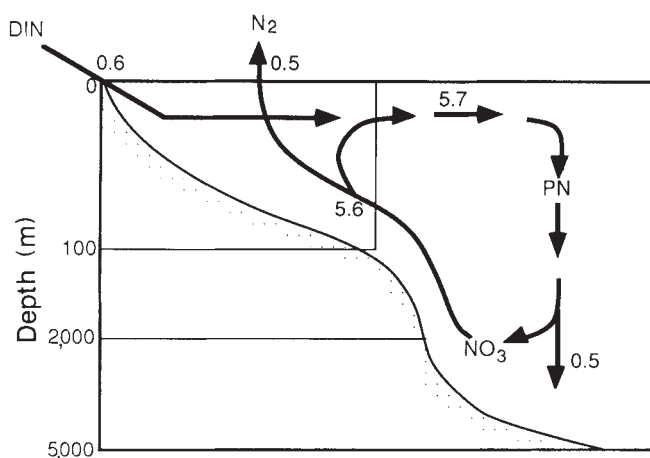


FIG. 1 Estimated fluxes ($10^{14} \text{ g N yr}^{-1}$) of riverine dissolved inorganic nitrogen (DIN), deep-sea nitrate (NO_3), particulate nitrogen (PN), and gaseous nitrogen (N_2) in the continental margins.

led nitrate suggest an f ratio (new/total production) of 0.54. In contrast, the same new production of 5.7×10^{14} g N yr⁻¹ over the deep sea (3.1×10^8 km²), a total primary production of 20×10^{15} g C yr⁻¹ here²⁴, and a carbon/nitrogen ratio of 5:1 suggest an f ratio of 0.14 for the oceanic ecosystem. These f ratios are consistent with tight coupling between primary producers, herbivores and saprovores in the deep sea, where large DOC excretion rates prevail. Here, export occurs mainly as fast-sinking (>100 m day⁻¹) faecal pellets, compared to phytodetrital export at similar settling velocities³³ from uncoupled shelf food webs.

Furthermore, a mean nitrate stock of $3.0 \mu\text{mol NO}_3 \text{ l}^{-1}$ at a depth of 100 m in the deep sea^{6,34} and a unit vertical influx of $1.84 \text{ g N m}^{-2} \text{ yr}^{-1}$ up the sharp oceanic gradient of nutrient (nutricline) yield an upwelling velocity of 43.8 m yr^{-1} beneath the euphotic zone (nominally 100 m). Over a similar 100-m interval of the nutricline, such an upwelling velocity is equivalent to a vertical eddy diffusivity K_z of $1.2 \text{ cm}^2 \text{ s}^{-1}$. Field studies of nitrogen uptake by phytoplankton and vertical gradients of nitrate indeed suggest an independent estimate of K_z of $\sim 1 \text{ cm}^2 \text{ s}^{-1}$ across the oceanic nutricline^{35,36}.

Marine phytoplankton are usually limited by the availability of nitrogen in the euphotic zone, not by the CO₂ concentration. Anthropogenic nitrogen loading has increased about tenfold in estuaries, where eutrophication is worldwide²⁴, for example, and it may be inducing phosphate limitation within the deep sea through rainfall²⁵. If future 'greenhouse' warmings lead to greater evaporation in the global hydrological cycle, such that the estuarine onwelling of nitrate increases as a consequence of larger freshwater runoff, we might expect another increment of the detrital carbon sink at the margins (provided that the annually averaged partial pressure of CO₂ in these waters is held below that of the atmosphere by the enhanced photosynthesis). Furthermore, coastal upwelling may increase under intensified along-shore wind stress associated with such a global warming³⁷, leading to additional carbon extraction from the atmosphere. How accurate, however, are the above onwelling and denitrification estimates? Extensive studies of continental margins, where half of the ocean biogeochemical fluxes may occur, are required. □

31. Jahnke, R. A., Reimers, C. E. & Craven, D. B. *Nature* **348**, 50–54 (1990).
32. Emerson, S. et al. *Nature* **329**, 51–53 (1987).
33. Alldredge, A. L. & Gotschalk, C. C. *Deep Sea Res.* **36**, 159–171 (1989).
34. Wroblewski, J. S., Sarmiento, J. L. & Flierl, G. R. *Global biogeochem. Cycles* **2**, 199–218 (1988).
35. Eppley, R. W., Renger, E. H. & Harrison, W. G. *Limnol. Oceanogr.* **24**, 483–494 (1979).
36. King, F. D. & Devol, A. H. *Limnol. Oceanogr.* **24**, 645–651 (1979).
37. Bakun, A. *Science* **247**, 198–201 (1990).
38. Dietrich, G. *General Oceanography Chart 4* (Wiley, New York, 1963).

ACKNOWLEDGEMENTS. This analysis was funded by the Department of Energy, NASA, the Office of Naval Research and the NSF.

Intermittent layered convection in a model mantle with an endothermic phase change at 670 km

Philippe Machetel & Patrice Weber

UPR 234/CNRS/GRGS 18 avenue E. Belin, 31055 Toulouse, France

IT is not yet known whether convection in the Earth's mantle occurs in cells that extend throughout the entire depth of the mantle, or if the upper and lower mantle convect in independent layers; both possibilities seem to be consistent with geochemical and seismological observations^{1–4}. It is generally accepted, however, that the seismologically observed boundary between the upper and lower mantle at 670 km depth arises from the endothermic phase change, spinel → perovskite + MgO. Here we investigate the results of including this phase change in a model of mantle convection. For realism and to eliminate scaling problems, our calculations use commonly accepted values of geophysical parameters. For a Clapeyron slope $\gamma = -2 \times 10^6$, a value close to experimental results and theoretical expectations, we observe local intermittent mixing between the upper and lower mantle. Such behaviour may offer a way to reconcile the existence of geophysical evidence for both whole-mantle and layered convection.

The effects of a phase change on the structure of convection have been studied previously in cartesian (planar) geometry for an incompressible fluid heated from below, by introducing a latent-heat release into the energy equation^{5,6}. The conditions of our model are, however, more relevant to the Earth's mantle: they include spherical, axisymmetric geometry, compressibility, internal heating, a phase change in the continuity equation and realistic geophysical values (Table 1) as input parameters⁷, instead of arbitrary dimensionless combinations.

Experimental and theoretical results have led to a consensus that the slope of the spinel/perovskite + MgO phase boundary is shallow and negative^{9–11}. With a Clapeyron slope $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$ and a transition depth of 670 km (ref. 12) at a temperature of 2,000 K (ref. 8), the phase boundary for our model is close to that inferred from experimental results^{9,10}. We have, however, studied the effect of varying the Clapeyron slope from $\gamma = 0$ to $\gamma = -2 \times 10^6$ to $\gamma = -4 \times 10^6 \text{ Pa K}^{-1}$.

The function $\Gamma(P, T)$ (see box on next page), which depends on the local pressure P and local temperature T , gives the ratio of the lower-mantle (perovskite + MgO) phase present in the mantle material. The reference density profile ρ_{ref} follows the Adams-Williamson equation above and below the phase change, allowing one to write an equation of state that takes into account the effects of pressure and temperature as perturbations. The continuity equation (see box on next page) takes into account the reference density and consequently the phase change as well. The latent heat release Q_1 is included in the energy equation. These equations, formulated in spherical, axisymmetrical geometry, were solved numerically with a finite-difference code that solves the vorticity and streamfunction equations. We checked that the equations of motion converged for each time-step (the lengths of which were controlled by the Courant criterion¹³). The calculations were run for 5,000 time-

Received 9 May 1990; accepted 14 January 1991.

1. Walsh, J. J., Dieterle, D. A. & Pribble, J. R. *Deep Sea Res.* (in the press).
2. Biscaye, P. E., Anderson, R. F. & Deck, R. L. *Cont. Shelf Res.* **8**, 885–904 (1988).
3. Jackson, G. A. et al. *Eos* **70**, 146–155 (1989).
4. Honjo, S., Manganini, S. J. & Cole, J. J. *Deep Sea Res.* **26**, 609–625 (1982).
5. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
6. Walsh, J. J. *Limnol. Oceanogr.* **21**, 1–13 (1976).
7. Bailey, G. W. & Chapman, P. in *South African Ocean Colour and Upwelling Experiment* (ed. Shannon, L. W.) 125–146 (Sea Fish. Res. Inst., Cape Town, 1985).
8. Walsh, J. J. et al. *Deep Sea Res.* **27**, 1–27 (1980).
9. McGill, D. A. in *The Biology of the Indian Ocean* (Ed. Zeitzschel, B.) 53–102 (Springer, Berlin, 1973).
10. Codispoti, L. A. & Friederich, G. E. *Deep Sea Res.* **25**, 751–770 (1978).
11. Smith, R. L. in *Coastal Upwelling* (ed. Richards, F. A.) 107–118 (American Geophysical Union, Washington, DC, 1980).
12. Walsh, J. J. *On the Nature of Continental Shelves* (Academic, New York, 1988).
13. Lee, T. N., Atkinson, L. P. & Legeckis, R. *Deep Sea Res.* **28**, 347–378 (1981).
14. Hogetsu, K. & Taga, N. in *Productivity of Biocenoses in Coastal Regimes of Japan* (eds Hogetsu, K., Hatanaka, M., Hanaoka, T. & Kawamura, T.) 31–172 (University of Tokyo, 1977).
15. Walsh, J. J., Dieterle, D. A., Meyers, M. B. & Muller-Karger, F. E. *Prog. Oceanogr.* **23**, 245–301 (1989).
16. Eismann, D. & Van Bennekom, A. J. *Hydrol. Newsl. Spec. Publ.* **6**, 25–48 (1971).
17. Lee, T. N. & Pietrafesa, L. J. *Prog. Oceanogr.* **19**, 267–312 (1987).
18. Brookman, U., Billen, G. & Gieskes, W. W. in *Pollution of the North Sea* (eds Salomons, W., Bayne, B. L., Duursma, E. K. & Forstner, U.) 348–389 (Springer, Berlin, 1988).
19. Takahashi, M. & Hori, T. *Mar. Biol.* **79**, 177–186 (1984).
20. Aagaard, K., Pease, C. H. & Salo, S. A. NOAA Tech. Mem. ERL PMEL-82, 1–171 (1988).
21. Walsh, J. J. et al. *Limnol. Oceanogr.* **23**, 659–683 (1978).
22. Beardsley, R. C., Chapman, D. C., Brink, K. H., Ramp, S. R. & Schlitz, R. J. *phys. Oceanogr.* **15**, 713–771 (1985).
23. Richey, J. E., Brock, J. T., Naiman, R. J., Wissmar, R. C. & Stallard, R. F. *Science* **207**, 1348–1351 (1980).
24. Walsh, J. J., Rowe, G. T., Iverson, R. L. & McRoy, C. P. *Nature* **291**, 196–201 (1981).
25. Fanning, K. A. *Nature* **339**, 460–463 (1989).
26. Knap, A., Jickells, T., Pszeny, A. & Galloway, J. *Nature* **320**, 158–160 (1986).
27. Walsh, J. J. *Bioscience* **34**, 499–507 (1984).
28. Christensen, J. P., Murray, J. W., Devol, A. H. & Codispoti, L. A. *Global biogeochem. Cycles* **1**, 97–116 (1987).
29. Devol, A. H. *Nature* **349**, 319–321 (1991).
30. Suzuki, Y. & Tanoue, E. in *Ocean Margin Processes in Global Change* (eds Mantoura, R. F., Martin, J. M. & Wollast, R. F.) (Wiley, New York, in the press).

TABLE 1 Notation and values for parameters

		Units	Value	Ref.
Acceleration of gravity	g	m s^{-2}	10.0	
Thermal expansion	α	K^{-1}	1.4×10^{-5}	20
Under-plate temperature	T_0	K	1,400	20
Core-mantle boundary temperature	T_i	K	3,000	4
External radius	r_0	m	6.291×10^6	12
Layer thickness	d	m	2.811×10^6	12
Heat capacity	C_p	$\text{J kg}^{-1} \text{K}^{-1}$	1.25×10^3	20
Thermal diffusivity	k	$\text{m}^2 \text{s}^{-1}$	0.8×10^{-6}	21
Dynamic viscosity	η	$\text{kg m}^{-1} \text{s}^{-1}$	10^{22}	22
Internal heating rate	H	$\text{J kg}^{-1} \text{s}^{-1}$	2.65×10^{-12}	23
Surface density	ρ_0	kg m^{-3}	3.37×10^3	12
Mid-depth density	ρ_m	kg m^{-3}	4.85×10^3	12
Bulk compressibility	K_s	Pa	4.3×10^{11}	12
Density jump (670 km)	$\delta\rho_{670}$	kg m^{-3}	390.0	12
Mean density (670 km)	ρ_{670}	kg m^{-3}	4.18×10^3	12
670-km pressure	P_{670}	Pa	2.383×10^{10}	12
670-km temperature	T_{670}	K	2,000	8
Thickness of transition	e	m	10^4	

steps, corresponding to time periods of 3.8×10^9 to 9.3×10^9 yr. The main computations were performed with a 513×81 -point grid, but tests of accuracy were run by comparing the chaotic evolution of the results with those from a $2,049 \times 321$ grid for $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$.

The local variation of the stream function between two points gives the amount of material flowing per unit time through the two-dimensional surface delimited by these points. Thus, zero contours of streamfunction define the boundaries of the convective cells, and within these cells the sign of the streamfunction indicates the rotation direction. The global structure of the streamfunctions are shown for $\gamma = 0$, -2×10^6 and $-4 \times 10^6 \text{ Pa K}^{-1}$ in the left hemispheres of Fig. 1, and the temperature fields are shown in the right. For $\gamma = 0$ (Fig. 1a), the convection structure is not affected by the phase change: cells can cross the phase boundary without distortion. In this case, the streamfunction is similar to that expected from chaotic, time-dependent convection: instabilities start from the top or bottom of the layer and join (or pair) with other co-rotating cells¹³. The temperature field above and below the discontinuity is not affected by the phase change because there is no latent-heat release. The presence of a large amount of internal heating favours cylindrical and sheet-like sinking currents with broad rising return flows¹⁴⁻¹⁶. For $\gamma = 0$, the sinking instabilities cross the phase boundary at a less-than-vertical dipping angle, directly below the point at which they are formed. On the other hand, a Clapeyron slope of $\gamma = -4 \times 10^6 \text{ Pa K}^{-1}$ hinders the penetration

EQUATIONS USED IN CALCULATIONS

Local relative fraction of lower-mantle phase:

$$\Gamma = \frac{1}{2} \left[1 + \tanh \left(\frac{P - P_{670} - \gamma(T - T_{670})}{\rho_{670} g e} \right) \right] \quad (1)$$

Reference density profile:

$$\rho_{\text{ref}} = \rho_a + \delta\rho_{670} \Gamma(P_H, T_{670}) \quad (2)$$

Adams-Williamson density profile:

$$\rho_a = \rho_0 \exp \left(\frac{g \rho_m (r_0 - r)}{K_s} \right) \quad (3)$$

Equation of state:

$$\rho = \rho_{\text{ref}} \left[1 - \frac{\rho_a}{\rho_{\text{ref}}} (\alpha(T - T_s) + K_T^{-1}(P - P_H)) + \frac{2\delta\rho_{670}}{\rho_{\text{ref}}} \frac{(P - P_H - \gamma(T - T_{670}))}{\rho_{670} g e} \times \Gamma(P_H, T_{670})(1 - \Gamma(P_H, T_{670})) \right] \quad (4)$$

Continuity equation:

$$\text{div}(\rho_{\text{ref}} \mathbf{v}) = 0 \quad (5)$$

Motion equation:

$$\rho \mathbf{g} + \text{div} \eta (\nabla \mathbf{v} + \nabla \mathbf{v}^T - \frac{2}{3} (\nabla \cdot \mathbf{v}) \mathbf{1}) - \nabla(P) = 0 \quad (6)$$

Latent-heat release:

$$Q_l = \frac{\gamma T \delta\rho_{670}}{\rho_{670}^2} \quad (7)$$

Energy equation:

$$\rho_{\text{ref}} \left[C_p \frac{D(T - T_s)}{Dt} \right] = \tau : \nabla \mathbf{v} + \rho_0 C_p k \Delta^2 T + \rho_{\text{ref}} H + \rho_{\text{ref}} Q_l \frac{D\Gamma}{Dt} \quad (8)$$

of sinking currents into the lower mantle (see the zero contour (thick line) of the streamfunction in Fig. 1c). Convection between the upper and lower mantle is now layered. The aspect ratio of the cells fluctuates, with cell widths of $\sim 1,000$ km in the upper mantle and $\sim 2,000$ km in the lower mantle. This discrepancy prevents perfect viscous coupling between cells in the lower and upper mantle. Viscously coupled cells should be distinguished by having streamfunctions of opposite sign, as rising currents in the lower cells would induce sinking currents in the upper. Except in a few locations (around the co-latitudes $\pi/4$ and $3\pi/4$), most of the cells adjust to allow for this kind

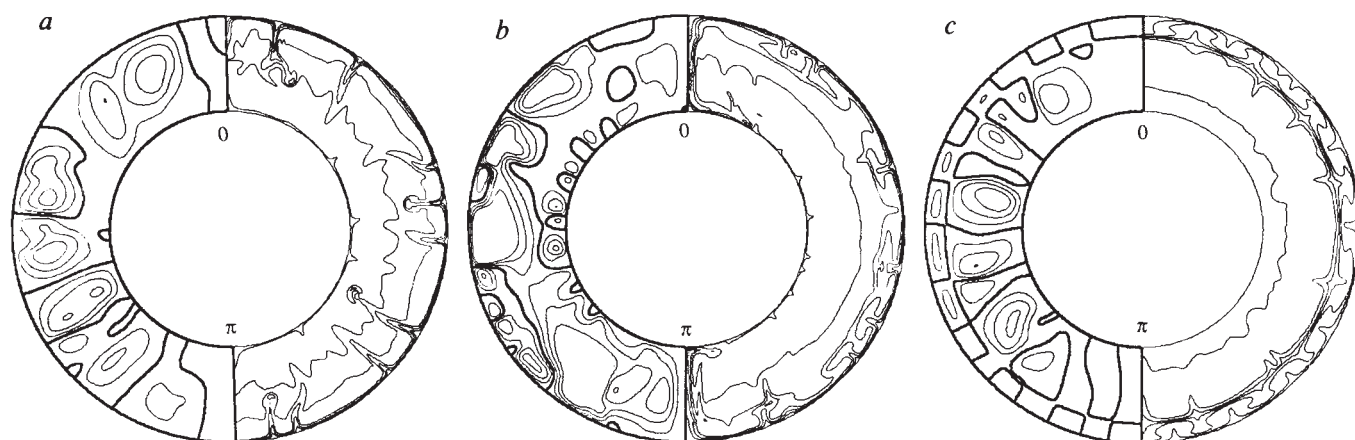
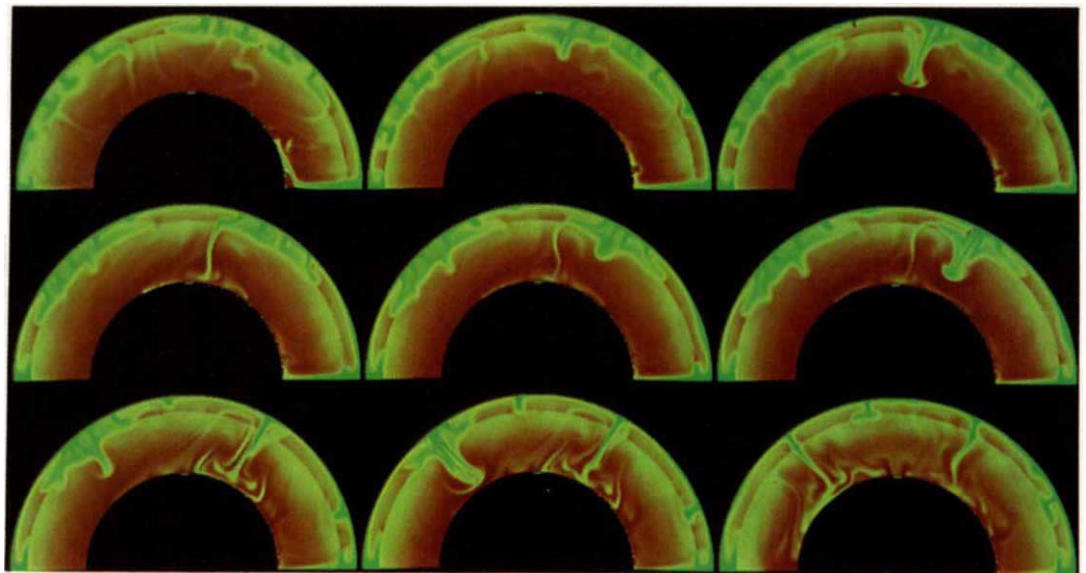


FIG. 1 Isocontours of streamfunctions and temperature fields for three values of the Clapeyron slope γ . No influence of the phase change is evident for $\gamma = 0$ (a), whereas a layered structure appears for $\gamma = -4 \times 10^6 \text{ Pa K}^{-1}$

(c) For $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$ (b), there is a mixture of whole-mantle and layered convection.

FIG. 2 Evolution of the thermal field for $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$. The time increases from left to right and from top to bottom. The cold sinking currents, individually unable to cross the phase boundary, penetrate into the lower mantle when they are pushed together by the rising return flows. When all the surrounding cold material has been drawn down penetration ceases, and spreading of new isolated instabilities at the discontinuity resumes.



of coupling. A Clapeyron slope $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$ is within the error ranges of experimental data^{8,9,10} and theoretical estimates ($-1.25 \times 10^6 \leq \gamma \leq -2.5 \times 10^6 \text{ Pa K}^{-1}$) (ref. 11). For this value, a mixed scenario of whole and layered convection appears. The streamfunction and the temperature field (Fig. 1b) show that convection is layered almost everywhere except at a few places where upper cells can penetrate into the lower mantle. It is difficult to establish what kind of coupling prevails. This structure, with local penetration of sinking currents and both thermal and viscous coupling, can explain why subducting slabs seem to be able either to bend at or to penetrate through the phase boundary³. It can also provide an explanation for the correlations and anticorrelations observed in the seismic-velocity anomalies above and below the 670-km discontinuity¹⁷.

Figure 2 illustrates the evolution over time of the temperature field for $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$. The sinking instabilities are stopped almost everywhere by the phase boundary, where they bend and spread. Some flows, however, are able to penetrate into the lower mantle locally. These intermittencies are caused by a chain of events. (1) The flow is dominated by sinking currents and broad return flows. The stronger upwellings push the cold anomalies towards each other. (2) The latter stick together until a threshold size is reached for which penetration into the lower mantle is possible. (3) Intermittent mixing of the upper and lower mantles ensues. All the surrounding cold anomalies are drawn in and sink until they reach the core. (4) The process stops when most of the cold material has disappeared. Then new, isolated sinking cells spread again on the discontinuity. These sequences occur both at the poles and at lower latitudes on the sphere, indicating that they can exist for both cylindrical and sheet-like structures.

An endothermic phase change induces a thermal boundary layer at the discontinuity^{3,4}. This temperature jump reaches 750 K for $\gamma = -4 \times 10^6 \text{ Pa K}^{-1}$ but only a few hundred degrees for $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$. For $\gamma = 0$ or $-4 \times 10^6 \text{ Pa K}^{-1}$ the mean temperature profiles are adiabatic below and above the discontinuity, whereas with $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$ the sudden intermittencies make the profiles sub-adiabatic at the bottom of both the lower and upper mantle. We calculated the time-averaged transfer of mass from the lower to the upper mantle during each run. Over 10^9 yr, 2% of the mantle crosses the discontinuity for $\gamma = -4 \times 10^6 \text{ Pa K}^{-1}$. For $\gamma = 0$, this rate fluctuates between 40% and 60% per 10^9 yr, with a time-averaged rate of 47%. For $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$, this rate is minimal ($\sim 10\%$) while layering exists, but peaks at $\sim 80\%$ for short periods during the mixing events, which occur roughly every 5×10^8 yr. The resulting time-averaged transfer rate is 22% over 10^9 yr.

Although they incorporate several simplifications, the mantle-

convection models presented here take into account the main gross properties of the real Earth (such as realistic core-mantle boundary and surface temperatures, mantle viscosity and compressibility, and internal heating). We have chosen to use a direct approach rather than imposing arbitrary, dimensionless combinations of parameters, to avoid scaling problems and to retain the physical significance of the results. The results for a Clapeyron slope $\gamma = -2 \times 10^6 \text{ Pa K}^{-1}$ are particularly interesting as this value is within the uncertainty range of both theoretical and experimental studies. Previous results⁶, obtained using a dimensionless approach in small-aspect-ratio cartesian geometry, showed layered convection or leaky layering only for much more endothermic Clapeyron slopes. The apparent discrepancy between these results can be attributed to our choice of a relatively low (but reasonable) value of α —a higher value would act to suppress the layering. The lateral boundary conditions imposed in cartesian geometry can also prevent the sinking currents from aggregating, however, and could thus suppress the occurrence of intermittent instabilities at the phase boundary. The intermittent penetrative convection events seen in our calculations, and the corresponding mass exchange rates, are compatible with the timescale of continent dispersal¹⁸ and with the existence of separated mantle reservoirs. They might provide the basis for models complementary to those based on chemical buoyancy¹⁹. Intermittent layered convection resulting from an endothermic phase change at the 670-km discontinuity might provide the key for reconciling the existing geophysical evidence for both whole-mantle and layered-mantle convective structures. □

Received 3 September 1990; accepted 18 January 1991.

- Dupré, B. & Allègre, C. J. *Nature* **303**, 142–146 (1983).
- Allègre, C. J., Hart, S. R. & Minster, J. F. *Earth planet. Sci. Lett.* **66**, 177–190 (1983).
- Olson, P., Carlson, P. G. & Carlson, R. W. *Nature* **344**, 209–214 (1990).
- Jeanloz, R. & Morris, S. A. *Rev. Earth planet. Sci.* **14**, 377–415 (1986).
- Christensen, U. R. & Yuen, D. A. *J. geophys. Res.* **89**, 4389–4402 (1984).
- Christensen, U. R. & Yuen, D. A. *J. geophys. Res.* **90**, 10291–10300 (1985).
- Machetel, P. *Geophys. Res. Lett.* **17**, 1145–1148 (1990).
- Boehler, R. & Chopelas, A. *Geophys. Res. Lett.* (submitted).
- Ito, E. & Katsura, T. *Geophys. Res. Lett.* **16**, 425–428 (1989).
- Ito, E. & Takahashi, E. *J. geophys. Res.* **94**, 10637–10646 (1989).
- Price, G. D., Wall, A. & Parker, S. C. *Phil. Trans. R. Soc. A* **328**, 391–407 (1989).
- Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
- Machetel, P. & Yuen, D. A. *J. geophys. Res.* **94**, 10609–10626 (1989).
- Houseman, G. *Nature* **332**, 346–349 (1988).
- Glatzmaier, G. A., Schubert, G. & Bercowski, D. *Nature* **347**, 274–277 (1990).
- Weinstein, S. & Olson, P. *Geophys. Res. Lett.* **17**, 239–242 (1990).
- Dziewonski, A. M. *J. geophys. Res.* **89**, 5929–5952 (1984).
- Le Pichon, X. & Huchon, P. *Earth planet. Sci. Lett.* **67**, 123–135 (1984).
- Ringwood, A. E. & Irfune, T. *Nature* **331**, 131–136 (1988).
- Stacey, F. D. *Phys. Earth planet. Inter.* **15**, 341–348 (1977).
- Anderson, D. L. *Theory of the Earth*, 366 (Blackwell Scientific, Boston, 1989).
- Hager, B. H. & Richards, M. A. *Phil. Trans. R. Soc. A* **328**, 309–327 (1989).
- Stacey, F. D. & Loper, D. E. *Phys. Earth planet. Inter.* **53**, 167–174 (1988).

What does it cost a plant to produce floral nectar?

Graham H. Pyke

Division of Environmental Science, The Australian Museum,
6–8 College Street, Sydney, New South Wales 2000, Australia

To understand the adaptive nature of floral nectar production it is necessary to determine for individual plants the associated costs and benefits in terms of growth and/or reproduction^{1–3}. Nectar production may use up to 37% of a plant's available energy^{4,5} but might not affect growth or reproduction. I report here that removal of nectar from flowers of Christmas bells (*Blandfordia nobilis*) increased the plant's net nectar production but reduced its ability to produce seeds. To our knowledge this is the first demonstration that nectar production entails a cost to a plant in terms of growth and/or reproduction and that both the gains and costs associated with nectar production may be estimated in the same 'currency' (seeds). As a plant's nectar production increases, there should therefore be a trade-off between pollinator-mediated increases in numbers of fertilized seeds^{1–3} and decreases in seed number due to the costs of producing the nectar.

The study was conducted during December 1988 and January 1989 in low heath in Barren Grounds Nature Reserve near Sydney (600 m above sea level, 34° 40' 45" S, 150° 42' 30" E). Here *Blandfordia nobilis* often occurs in large dense patches. The plant is an erect perennial herb, having three to ten flowers, and being visited by nectar-feeding honey-eaters (Aves: Meliphagidae) and by pollen-collecting honey bees (*Apis mellifera*)^{6–8}.

The following experiment demonstrates that, as in some other plant species (for example *Echium vulgare*⁹; *Heliconia imbricata*¹⁰), removal of nectar from *Blandfordia* flowers increases net production of nectar. During January 1989, 50 plants were selected on which at least one flower had opened on the first morning of the experiment. For each plant, one such flower was chosen and the plant bagged with mosquito netting to exclude visits by all animals except ants. To exclude ants a coating of Tanglefoot (Tanglefoot Co.) was applied to the stem of each plant, and all vegetation that made contact with the plant was removed. The plants were then numbered in the sequence in which they were encountered. On the first three mornings the selected flower on each odd numbered plant was probed without nectar removal using blocked capillary tubes, while from the selected flower on each even numbered plant nectar was removed with normal unblocked tubes. Sugar concentration was measured for each nectar sample using a Bellingham and Stanley refractometer adjusted by the manufacturer for small volumes and expressed (after transformation) in terms of weight of equivalent sucrose per unit volume¹¹. On the fourth morning, nectar was removed from all 50 flowers. The flowers

TABLE 1 Net sugar production and frequency of sampling

	Total net sugar* production (mean ± s.e.m. (sample size))
Daily nectar removal	23.3 ± 2.6 (25)
Nectar removal only on fourth day	8.8 ± 1.5 (25)
Probability†	<0.001

* Measured in mg sucrose equivalent.

† Based on two-tailed Student's *t*-test.

from which nectar had been removed each day yielded about three times as much nectar sugar as the flowers from which no nectar had been removed until the fourth day (Table 1).

We next measured the cost to a plant associated with nectar production. In both December and January, flowering plants were selected on which no flowers had opened, but on which at least one flower was expected, from its size and colour, to open the next day. The plants were numbered in the sequence in which they were encountered. Each day for the next three weeks (until 95% of flowers had opened) the plants numbered 1, 4, 7 and so on, had nectar removed from all open flowers, while all open flowers on the plants numbered 2, 5, 8 and so on, were probed without removing nectar. Each day all open flowers on each plant in the two groups were artificially cross-pollinated by rubbing anthers with a lot of pollen across the stigmatic surfaces. The anthers were taken from plants at least 10 m away from the recipient plant. Thus the number of seeds for each plant would have been limited only by available resources and not by the level of pollination. All plants in these two groups were also kept bagged with mosquito netting until all flowering had ceased. The plants numbered 3, 6, 9 and so on, were not treated in any way and were not bagged. Six weeks after the ends of the December and January experiments, respectively, all the plants were excavated, washed clean, frozen and taken back to the laboratory where they were separated into seeds, fruit capsule and surrounding bracts, other above-ground parts (stem and leaves) and below-ground parts (rhizome and roots). The various plant parts were oven-dried for 24 h at 40 °C and then weighed. The seeds from each fruit were also counted. Because small numbers of seeds could not be accurately weighed, total seed weights for plants with fewer than 10 seeds were not included in the analysis. Plants were also excluded if any of their flowers, fruits or seeds were eaten.

During January, seed number per plant was limited by the level of pollination as the average seed number for January was significantly lower for untreated plants than for plants which were artificially pollinated but from which no nectar had been removed (Table 2). Similar pollination-limited seed number has been found for each of the previous four years (ref. 11, and G.H.P., unpublished data). For December, the difference, though in the same direction, was not significant (Table 2). That

TABLE 2 Summary of results of nectar-removal and pollination experiment

	Untreated	Pollinated	Pollinated and nectar removed	Significance (see notes)
Seeds produced per plant	Dec. 121 ± 24 (16) Jan. 12 ± 7 (7) Both 149 ± 14 (27)	145 ± 17 (17) 156 ± 25 (10) 116 ± 16 (29)	1	
Weight per seed (mg)	8.99 ± 0.35 (15)	7.80 ± 0.40 (26)	7.84 ± 0.21 (23)	2
Weight of stem and leaves (g)	1.66 ± 0.22 (23)	1.26 ± 0.11 (25)	1.48 ± 0.14 (27)	2
Weight of rhizome and roots (g)	2.92 ± 0.38 (23)	1.83 ± 0.23 (25)	2.23 ± 0.40 (27)	2
Weight of fruit capsule and bracts (g)	0.32 ± 0.06 (23)	0.32 ± 0.02 (27)	0.31 ± 0.03 (26)	2
Number of flowers per plant	5.39 ± 0.60 (23)	4.52 ± 0.33 (27)	4.97 ± 0.41 (29)	2

(1) Month by treatment interaction was significant ($P < 0.05$) in two-way analysis of variance (ANOVA) with three treatments and two months as independent variables but not when control treatment was excluded. The number of seeds produced per plant was significantly greater for pollinated plants than for control plants in January ($P < 0.05$, one-way ANOVA) but not in December ($P > 0.05$, one-way ANOVA). The number of seeds produced per plant was significantly greater for pollinated plants than for plants that were pollinated and from which nectar had been removed ($P = 0.02$, two-way ANOVA with two treatments and two months as independent variables). (2) Differences among treatments and month by treatment interaction were not significant ($P > 0.05$, two-way ANOVA with three treatments and two months as independent variables).

seed number per plant is limited by pollination has rarely been demonstrated (see ref. 12).

The increase in the net production of nectar by the plants from which nectar had been removed entailed a cost to them: their seed number was significantly less than that of plants from which no nectar had been removed and which therefore had a lower net production of nectar (Table 2). No costs of nectar production were apparent in terms of the other aspects of plant growth or reproduction (Table 2). For plants of this species, both gains and costs associated with increased (or decreased)

nectar production can therefore be measured in the same units (seeds). All else being equal, a plant that produces more nectar will have higher average nectar-standing crops per flower when visited by a pollinator^{1,2,13,14}, and will therefore have on average more flowers probed during each visit by a pollinator, receive and transmit more pollen, and produce more seeds (its own and those of other plants that it has fathered)^{1,2,13}. There should, therefore, be a trade-off between extra fertilized seeds resulting from enhanced nectar production and the cost in terms of seeds of producing the nectar. □

Received 23 July; accepted 31 December 1990.

1. Pyke, G. H. *Theor. Populat. Biol.* **20**, 326–343 (1981).
2. Pyke, G. H. in *Foraging Behavior, Ecological, Ethological, and Psychological Approaches* (eds Kamil, A. C. & Sargent, T. D.) (Garland, New York, 1981).
3. Pyke, G. H. & Waser, N. M. *Biotropica* **13**, 260–270 (1981).
4. Pleasant, J. M. & Chaplin, S. J. *Oecologia* **59**, 232–238 (1983).
5. Southwick, E. E. *Ecology* **65**, 1775–1779 (1984).
6. Pyke, G. H., Day, L. P. & Wale, K. A. *Aust. J. Ecol.* **13**, 279–284 (1988).
7. Zimmerman, M. & Pyke, G. H. *Aust. J. Ecol.* **13**, 93–99 (1988).
8. Zimmerman, M. & Pyke, G. H. *Aust. J. Ecol.* **13**, 301–309 (1988).

9. Corbet, S. A. & Delfosse, E. S. *Aust. J. Ecol.* **9**, 125–139 (1984).
10. Gill, F. B. *Biotropica* **20**, 169–171 (1988).
11. Pleasants, J. M. *Am. J. Bot.* **70**, 1468–1475 (1983).
12. Zimmerman, M. & Pyke, G. H. *J. Ecol.* **76**, 777–789 (1988).
13. Zimmerman, M. *Oikos* **41**, 57–63 (1983).
14. Zimmerman, M. *Am. Midl. Nat.* **120**, 50–57 (1988).

ACKNOWLEDGEMENTS. I thank S. Corbet, A. Jones, D. Paton, E. Southwick, N. Waser and M. Zimmerman for comments on the manuscript, and D. Ayers, G. Barrett, M. Christy, B. Draper, G. Gowing and S. Frazer for assistance with field work.

Light-induced suppression of endogenous circadian amplitude in humans

Megan E. Jewett^{*†}, Richard E. Kronauer[‡] & Charles A. Czeisler^{*§}

^{*} Laboratory for Circadian and Sleep Disorders Medicine, Department of Medicine, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA
[‡] Division of Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

WINFREE reported 20 years ago the intriguing finding that a light stimulus of a critical strength applied at a critical circadian phase could essentially stop the circadian clock in *Drosophila pseudoobscura* by resetting the circadian oscillator close to its singularity (a phaseless position at which the amplitude of circadian oscillation is zero)¹. Since then, similar observations of attenuated circadian amplitude in response to critical stimuli have been limited to unicells, insects and plants^{2–7}. Our recent demonstration that the phase of the human circadian pacemaker could be inverted using an unconventional three-cycle stimulus^{8,9} led us to investigate whether critically timed exposure to a more moderate stimulus could drive that oscillator towards its singularity. Here we report that exposure of humans to fewer cycles of bright light, centred around the time at which the human circadian pacemaker is most sensitive to light-induced phase shifts, can markedly attenuate endogenous circadian amplitude. In some cases this results in an apparent loss of rhythmicity, as expected to occur in the region of singularity.

When a resetting stimulus is so weak that it can reset the phase of a circadian oscillator by only a small amount (Fig. 1a), a plot of the time of stimulus application with respect to two different timescales—the initial (pre-stimulus) phase of the circadian oscillation versus the final (post-stimulus) phase of that oscillation—has an average slope of one (Fig. 1b). Hence such weak resetting is labelled type 1 (ref. 1). On the other hand, if the strength (*S*) (duration and intensity) of a stimulus exceeds a critical strength (*S**), the stimulus can reset the phase of a circadian oscillator by large amounts of up to 12 hours (Fig.

1a). In this case, a similar plot of the stimulus time with respect to the initial versus the final phase of the circadian oscillation has an average slope of zero (Fig. 1b). Hence such strong resetting is called type 0 (ref. 1). The difference between type 1 and type 0 resetting is especially apparent at a critical time (*T**) (which typically occurs near the steepest portion of the type 0 phase response curve (see Fig. 1)). Winfree found that when a stimulus of critical strength (*S**) (between that which induces type 1 and that which induces type 0 resetting) was applied at this critical time, *T**, the circadian oscillation could be driven close to its singularity, such that it displayed an amplitude near zero^{1,10}, rather than a predictable change in phase. Since then, investigations in diverse organisms have confirmed this finding and have shown that the responses of the circadian pacemaker to near-critical stimuli are highly sensitive to initial conditions^{1,3,5}, such that very small variations in the initial time (*T*) or in the effective strength (*S*) of a stimulus (*T, S*) result in large variations in both final phase⁶ and final amplitude^{4,10}. As (*T, S*) nears (*T*, S**), extreme reductions in amplitude^{7,10}, abnormalities in waveshape^{2,5,6} and/or large shifts to unpredictable phases⁵ may be observed. We hypothesized that the human circadian system would display a similar variety of responses after exposure to near-critical stimuli.

We carried out 18 experimental trials with 14 young male subjects (mean $\pm$ s.e.m. age, 22.3 ± 0.8 yr; range, 18–29 yr). Each trial began with an initial pre-stimulus evaluation of endogenous circadian phase and amplitude using a 30- to 50-h constant-routine technique. This procedure is designed to unmask the endogenous components of the rhythms of core body temperature and plasma cortisol concentration by either eliminating, or distributing uniformly across the circadian cycle, physiological responses to environmental and behavioural stimuli that can otherwise obscure them^{8,11}. This initial constant routine was followed by exposure to a near-critical stimulus intended to reduce circadian amplitude (either one 8–9-h bright light episode within one circadian cycle or two 5–6-h episodes within two successive circadian cycles). Endogenous circadian phase and amplitude were then reassessed by a final post-stimulus constant routine.

In each subject's initial constant routine, rhythms of temperature and cortisol showed a normal amplitude and a normal phase relationship to the subject's habitual sleep–wake schedule. The responses of these circadian markers to a near-critical stimulus fell into three groups (Table 1). In group I (three trials), there was a loss of circadian rhythmicity in both temperature and cortisol throughout the final constant routine (Fig. 2). In group II (seven trials), large shifts to unpredictable phases were

[†] Present address: Psychology Department, University of California at Berkeley, Berkeley, California 94720, USA.

[§] To whom correspondence should be addressed.

observed in both the temperature and cortisol rhythms, and the circadian amplitude of one or both of these variables was partially reduced (Fig. 3). Finally, in group III (five trials), temperature and cortisol each responded differently to the near-critical stimulus: circadian amplitude was markedly attenuated in one of the variables, while the other variable continued to cycle at a partially reduced amplitude, often with an abnormal waveshape and at a shifted phase (Fig. 4). In all cases, episodic secretion of cortisol was found to persist, even in the absence of a prominent 24-h drive (Fig. 2d), suggesting that the pulse-generator driving the episodic secretion of this hormone may be functionally distinct from the circadian pacemaker. In three additional trials for which cortisol data are not available, one subject showed a partial reduction in the amplitude of his temperature rhythm, and two subjects showed a near-complete loss of rhythmicity.

Both schedules of exposure to bright light generated similar results (see Table 1), indicating that their strengths are roughly equal. As expected, the responses of individual subjects to near-critical stimuli varied (see Table 1), in contrast to the

consistent responses of subjects exposed to non-critical stimuli at a given phase⁸. Because the exposure of the subjects to bright light at a critical phase required the inversion of their sleep-wake cycles, our results could be due to this inversion rather than to the light exposure. We think that this is unlikely, however, because in a series of eight similar experiments in which scheduled exposure to darkness was substituted for exposure to bright light for three cycles, the temperature rhythm showed no significant change in amplitude (mean value $\pm$ s.e.m. $q(t) = 0.96 \pm 0.06$; compare with Table 1) and an average change in phase of only 1.46 h (C.A.C. *et al.*, unpublished data).

In the three trials in which both temperature and cortisol were nearly arrhythmic (group I), the reduction in amplitude was maintained throughout the final constant routine without evidence of recovery of a 24-h rhythm (Fig. 2c and d). Similar studies in other organisms have reported a loss of rhythmicity which continues with no evidence of regrowth for days^{1,2,4-6,10} or even a week^{3,7} in cases of extreme amplitude reduction. In this study, however, practical considerations constrained the duration of the constant-routine procedure. Therefore, the observed reductions in circadian amplitude could reflect the output of an oscillator that has not yet reached its final phase position rather than one that is roughly stationary near its region of singularity. Our findings, however, are consistent with three important features observed in disparate species monitored for extended durations after exposure to near-critical stimuli: (1)

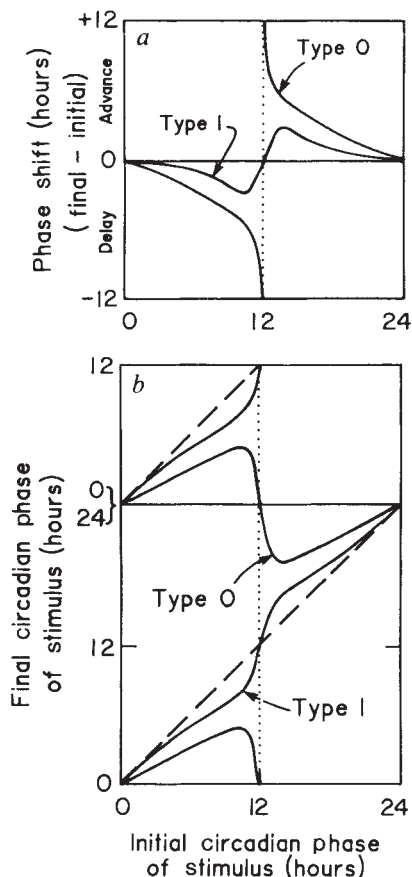


FIG. 1 Schematic illustrations of circadian phase resetting. *a*, Weak type 1 and strong type 0 phase response curves. The abscissa represents the phase at which the stimulus occurred with respect to the initial pre-stimulus circadian oscillation. The amount of phase shift induced by the stimulus is plotted on the ordinate. *b*, Weak type 1 and strong type 0 phase resetting curves. Abscissa as in *a*. The ordinate represents the phase at which the stimulus occurred with respect to the final, post-stimulus circadian oscillation. Note that in type 1 resetting the response to the stimulus is so weak that the phase resetting curve has an average slope of one and deviates only slightly from the dashed line of equality, which represents the complete absence of a resetting response. Note that in type 0 resetting, the response to the stimulus is so strong that the phase resetting curve has an average slope of zero and the phase relationship of the stimulus to the final circadian oscillation is restricted to a narrow band of phases. In both *a* and *b*, a vertical dotted line is plotted at the critical initial phase (T^*). A stimulus of critical strength (S^*) given at this time will induce a singular response.

TABLE 1 Responses of temperature and cortisol rhythms to near-critical stimuli

Group I: Substantial loss of rhythmicity in both variables

Subject	T	S	$q(t)$	$q(c)$	$\Delta\phi(t)$	$\sim\Delta\phi(c)$
642	+0.2	5.0×2	0.30	0.45	—	—
703 (trial 2)	+0.5	5.5×2	0.02	0.10	—	—
917	-0.4	5.6×2	0.24	0.38	—	—

Group II: Partial amplitude reduction and phase shift in both variables

Subject	T	S	$q(t)$	$q(c)$	$\Delta\phi(t)$	$\sim\Delta\phi(c)$
724 (trial 1)	+1.8	5.5×2	0.48	0.50	-4.9	+1
724 (trial 3)	+1.1	6.0×2	0.85	0.44	-9.9	-11
724 (trial 4)	+1.2	6.0×2	0.44	0.59	+6.8	+7
910	+0.6	6.2×2	1.22	0.53	+7.7	+8
911	-0.1	9.0×1	0.83	0.44	-3.3	-4
914	+0.7	6.2×2	0.36	0.97	+8.8	+7
922	+0.8	5.6×2	0.95	0.52	-4.7	-5

Group III: Substantial loss of rhythmicity in one variable; partial reduction and phase shift in the other variable

Subject	T	S	$q(t)$	$q(c)$	$\Delta\phi(t)$	$\sim\Delta\phi(c)$
831	+0.4	8.0×1	0.16	0.69	—	0
913	-0.1	9.0×1	0.32	0.84	—	+5
916	+0.7	6.2×2	0.31	0.82	—	+6
919	+0.6	5.6×2	0.28	0.80	—	+5
926	+0.6	5.6×2	0.59	0.32	+5.5	—

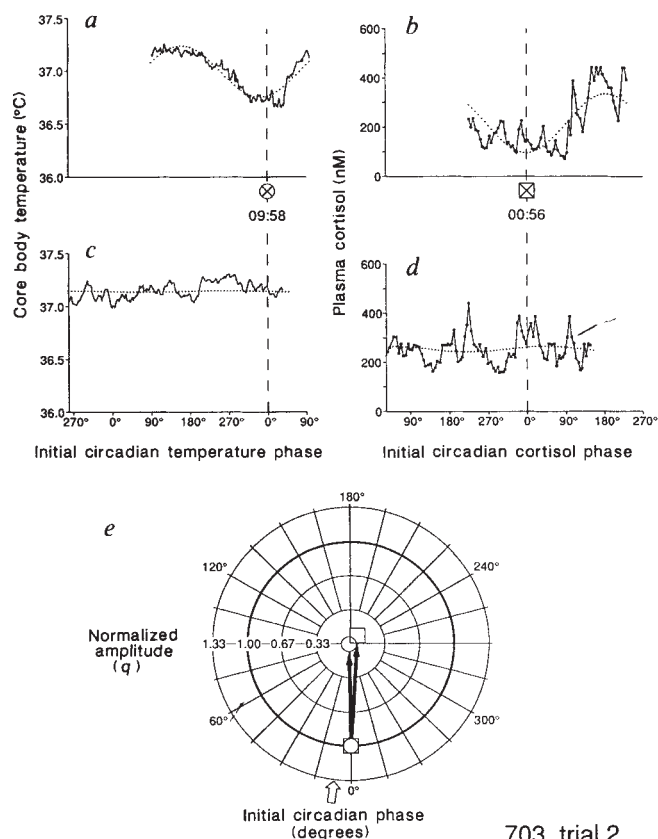
Trials in which cortisol data are not available:

Subject	T	S	$q(t)$	$\Delta\phi(t)$
703 (trial 1)	+0.2	5.4×2	0.60	-7.2
724 (trial 2)	+1.2	5.5×2	0.30	—
1027	0.0	5.5×2	0.15	—

T , time (hours before (-) or after (+) the initial fitted temperature minimum) at which the weighted midpoint of the total daily light exposure occurred; S , length of bright light exposure (hours) $\times$ number of cycles; q , (fitted amplitude in final constant routine)/(fitted amplitude in initial constant routine); $q(t)$, q for temperature data; $q(c)$, q for cortisol data. —, Amplitude too low to assess final phase position accurately. $\Delta\phi(t)$, Phase change in temperature data (hours); $\sim\Delta\phi(c)$, approximate phase change in cortisol data (hours).

FIG. 2 Endogenous circadian rhythms of core body temperature and plasma cortisol levels measured in an 18-year-old man (subject 703, trial 2) in group I. *a*, Dual harmonic regression model (dotted line) was fitted using nonlinear least squares¹⁶ to the temperature data (solid line) collected during the initial pre-stimulus constant routine to identify the fitted temperature minimum (encircled cross), which occurred at 09:58. That fitted minimum has been assigned a reference value of 0 degrees on the abscissa. The initial fitted amplitude of the single harmonic component of the model was 0.24 °C. *b*, Although cortisol is secreted episodically, this same model (dotted line) was used to estimate the amplitude and approximate the phase of the circadian component of the cortisol values (solid line) collected during the initial constant routine. The minimum of the single harmonic component of the model (boxed cross), which occurred at approximately 00:56, has been assigned a reference value of 0 degree on the abscissa. The amplitude of the fitted curve for cortisol was 119 nmol l⁻¹. During the post-stimulus constant routine the fitted temperature amplitude was reduced to 0.005 °C (*c*, symbols as in *a*), and the fitted cortisol amplitude was reduced to 11 nmol l⁻¹ (*d*, symbols as in *b*). *e*, Polar representation of the above data, in which circadian phase is expressed in degrees (of both the temperature and cortisol cycles) and the fractional reduction in amplitude (*q*) is indicated on the radius, with the initial amplitude of both the temperature rhythm (open circle) and cortisol rhythm (open box) normalized to a value of 1.00. Polar vectors (filled arrows) represent the changes in amplitude and phase of the oscillations. Note that near the origin, phase is indeterminant. The open arrow represents the circadian phase of the weighted midpoint of the total daily light exposure⁸ referenced to the temperature cycle.

METHODS. After providing informed consent, each subject was studied in an environment free of time cues and was initially scheduled to sleep at his habitual time. The initial constant routine was then begun upon awakening and carried out in ordinary indoor-room light (~150 lux), as described in detail elsewhere¹⁷. Subjects were exposed to a cyclical stimulus of bright light (7,000–12,000 lux), ordinary indoor-room light, and darkness, as described in ref. 8, with the weighted midpoint of the total daily light exposure centred near the initial fitted temperature minimum, which corresponds to the steepest portion of the type 0 phase resetting curve reported previously⁸. Using a mathematical model developed by Kronauer¹⁸ to predict *S**, two alternative schedules for delivery of the bright-light component of the stimulus were derived: (1) exposure to a single long (8–9 h) episode of bright light within about one circadian cycle (three trials); or (2) exposure to two shorter (5–6.2 h) episodes of bright light, each on successive days, within about two circadian cycles (15 trials). Sleep–dark episodes were centred roughly 12 h opposite the midpoint of the bright-light episodes, and thus occurred during the subjects' normal waking hours.



703, trial 2

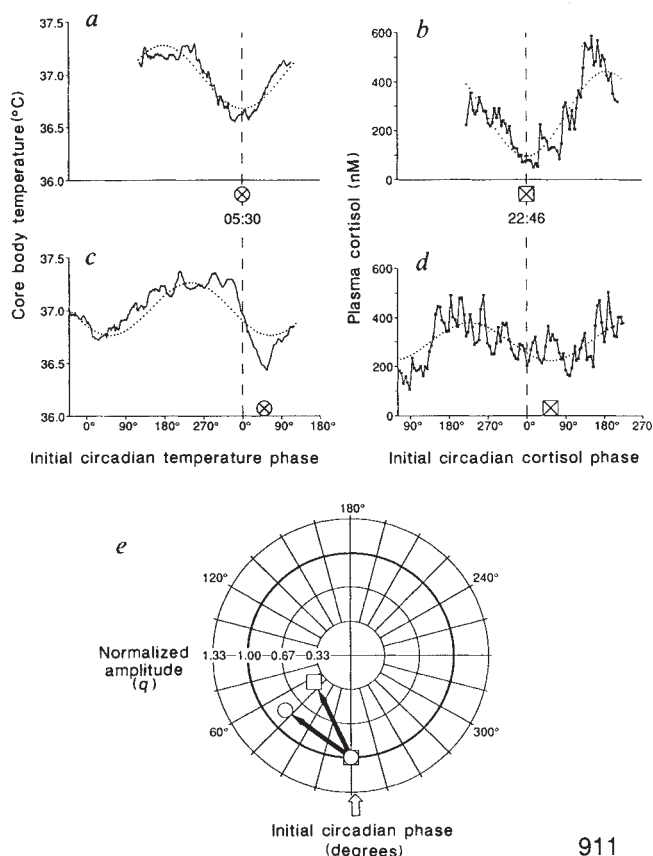


FIG. 3 Endogenous circadian temperature and cortisol rhythms in a 19-year-old man (subject 911) in group II. Symbols as in Fig. 2. *a*, Initial fitted temperature minimum occurred at 05:30 and the initial fitted temperature amplitude was 0.30 °C. *b*, Initial fitted cortisol minimum occurred at about 22:46 and the initial fitted cortisol amplitude was 174 nmol l⁻¹. During the post-stimulus constant routine the fitted temperature amplitude was reduced to 0.25 °C and the fitted temperature minimum occurred at 08:49, which represents a phase delay of 3.3 h (*c*), and the fitted cortisol amplitude was reduced to 77 nmol l⁻¹ and the fitted cortisol minimum occurred at about 02:22, which represents a phase delay of about 4 h (*d*). *e*, Above data are summarized on polar coordinates.

911

T^* occurs at the steepest portion of the type 0 phase resetting curve for human subjects⁸; (2) S^* is of a duration intermediate between that which induces type 1 and that which induces type 0 resetting; and (3) near-critical stimuli induce either a loss of circadian rhythmicity with no evidence of growth in circadian amplitude (as seen in group I), or large shifts to unpredictable phases (as seen in group II). This consistency with previous reports suggests to us that the amplitude attenuation seen in group I does indeed reflect a pacemaker near its region of singularity.

The surprising results for group III, where one marker showed a loss of circadian rhythmicity while the other continued to oscillate, cannot be compared with results seen in the singular region of other organisms because previous studies have monitored only one circadian variable. It is possible that even after some circadian variables have reached their final amplitudes and phase positions, other variables are still in transition. Alternatively, variables could differ in sensitivity to pacemaker output, and after exposure to an amplitude-reducing stimulus, the output of the pacemaker may be too weak to induce a circadian rhythm in one variable, but sufficient to induce rhythmicity in another. The group III results nevertheless suggest that the temperature and cortisol rhythms may have independent links to the circadian pacemaker.

That the singular region seems to be attainable in a mammalian vertebrate such as the human supports Winfree's contention that the existence and characteristics of the singular region

of an oscillator are based on a mathematical construct—the topology of the resetting surface—rather than on the anatomical organization of the circadian pacemaker¹². Previous data indicating that it is possible to achieve type 0 resetting in other vertebrates^{13,15} have led to the prediction of a singular region in vertebrate circadian systems^{12,14}. Our results are consistent with this hypothesis, and further emphasize the importance of measuring the amplitude as well as the phase of endogenous circadian rhythms¹⁰. Further exploration of the singular region in humans and other mammals may aid in the development and testing of mathematical models of pacemaker function, as well as contribute to the practical application of chronobiology to the circadian aspects of jet-lag, shift work and age-related attenuation of circadian amplitude. □

Received 4 September; accepted 14 December 1990.

- Winfree, A. T. *J. theor. Biol.* **28**, 327–374 (1970).
- Saunders, D. S. *J. comp. Physiol.* **124**, 75–95 (1978).
- Engelmann, W. & Johnsson, A. *Physiol. Plant* **43**, 68–78 (1978).
- Walz, B. & Sweeney, B. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6443–6447 (1979).
- Peterson, E. L. *J. comp. Physiol.* **138**, 201–211 (1980).
- Taylor, W., Krasnow, R., Dunlap, J. C., Broda, H. & Hastings, J. W. *J. comp. Physiol.* **148**, 11–25 (1982).
- Malinowski, J. R., Laval-Martin, D. L. & Edmunds, L. N. *J. comp. Physiol.* **155**, 257–267 (1985).
- Czeisler, C. A. *et al. Science* **244**, 1328–1333 (1989).
- Strogatz, S. H. *J. Biol. Rhythms* **5**, 169–174 (1990).
- Winfree, A. T. *J. comp. Physiol.* **85**, 105–140 (1973).
- Mills, J. N., Minors, D. S. & Waterhouse, J. M. *J. Physiol., Lond.* **285**, 455–470 (1978).
- Winfree, A. T. *The Timing of Biological Clocks* (Scientific American Books, New York, 1987).
- Eskin, A. in *Biochronometry* (ed. Menaker, M.) 55–80 (National Academy of Sciences, Washington, DC, 1971).
- Enright, J. T. & Winfree, A. T. *Lect. Math. Life Sci.* **19**, 121–150 (1986).
- Gander, P. H. & Lewis, R. D. *Am. J. Physiol.* **245**, R10–R17 (1983).
- Brown, E. N. thesis, 1–116, Harvard Univ. (1987).
- Czeisler, C. A. *et al. New Engl. J. Med.* **322**, 1253–1259 (1990).
- Kronauer, R. E. in *Sleep '90* (ed. Horne, J. A.) 306–309 (Pontenagel, Bochum, 1990).

ACKNOWLEDGEMENTS. We thank the subject volunteers, the student research technicians, J. S. Allan, E. N. Brown, J. F. Duffy, S. C. Hinton, B. C. Potter, G. S. Richardson, J. M. Ronda and T. L. Shanahan for assistance, and the US National Aeronautics and Space Administration and the National Institutes on Aging and of Mental Health for their support.

Enhancement of T-cell responsiveness by the lymphocyte-specific tyrosine protein kinase $p56^{lck}$

Ninan Abraham*, M. Carrie Miceli†, Jane R. Parnes† & André Veillette*‡§

* McGill Cancer Centre and ‡ Department of Biochemistry, McGill University, Montréal, Québec H3G 1Y6, Canada
† Division of Immunology and Rheumatology, Department of Medicine, Stanford University Medical Center, Stanford, California 94305, USA

LYMPHOCYTE-specific tyrosine protein kinase $p56^{lck}$ is physically associated with CD4 and CD8 T-cell surface molecules, suggesting that it may transduce CD4/CD8-triggered tyrosine phosphorylation signals during antigen stimulation^{1,2}. Indeed, antibody-mediated aggregation of CD4 (to mimic interaction with its ligand, major histocompatibility complex (MHC) class II molecules), rapidly elevates the kinase activity of $p56^{lck}$ (refs 3, 4) and is associated with marked changes in tyrosine protein phosphorylation^{4–6}. Genetic analyses suggest that the interaction of CD4/CD8 with $p56^{lck}$ results in a positive signal during antigen-induced T-cell activation^{7–10}. To evaluate directly the role of $p56^{lck}$ in T-cell activation, we introduced a constitutively activated form of Lck protein (tyrosine 505 to phenylalanine 505 mutant; refs 11–13) in a CD4-negative, MHC-class II restricted mouse T-cell hybridoma¹⁴. We report here that, as for transfection of CD4^{10,15,16}, expression of the Lck mutant enhanced T-lymphocyte

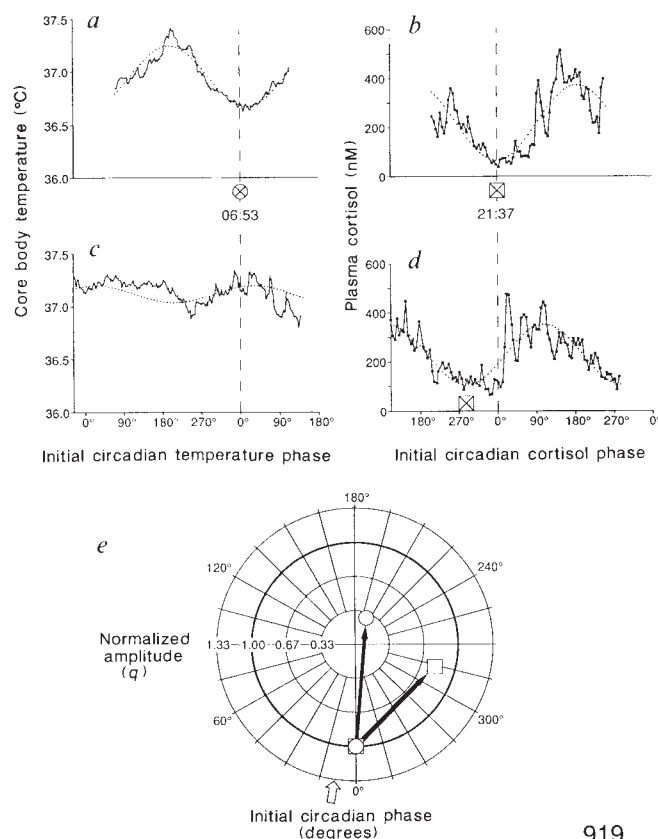


FIG. 4 Endogenous circadian temperature and cortisol rhythms in a 29-year-old man (subject 919) in group III. Symbols as in Fig. 2. a, Initial fitted temperature minimum occurred at 06:53 and the initial fitted temperature amplitude was 0.28 °C. b, Initial fitted cortisol minimum occurred at about 21:37 and the initial fitted cortisol amplitude was 155 nmol l⁻¹. During the post-stimulus constant routine, the fitted temperature amplitude was reduced to 0.08 °C (c); and the fitted cortisol amplitude was reduced to 121 nmol l⁻¹ and the fitted cortisol minimum occurred at about 16:42, which represents a phase advance of about 5 h (d). e, Above data are summarized on polar coordinates.

§ To whom correspondence should be addressed.

responsiveness. This finding provides direct evidence that $p56^{lck}$ can positively regulate T-cell functions and that it mediates at least some of the effects of CD4 and CD8 on T-cell activation.

BI-141 is a CD4-negative, class II MHC-restricted mouse helper T-cell hybridoma specific for beef insulin¹⁴. BI-141 cells were infected with retroviruses expressing both a mutant $p56^{lck}$ polypeptide having constitutively elevated tyrosine protein kinase activity, and neomycin phosphotransferase (Fig. 1a). This Lck mutant (F505 $p56^{lck}$) carries a tyrosine-to-phenylalanine substitution at tyrosine 505, the major site of Lck tyrosine phosphorylation *in vivo* (refs 11–13). Control cells were generated by infection with retroviruses encoding neomycin phosphotransferase alone. Five of the mutant-expressing cell lines were randomly selected for further analysis (Fig. 1b). These contained 5–11 times the levels of Lck protein (lanes 6–10) of control G418-resistant cell lines (Neo) (lanes 1–5). Cytofluorometric analyses showed that all cells were CD4-negative and expressed levels of T-cell receptor (TCR), CD3, Thy-1.2 and CD45 comparable with those of parent BI-141 and control cells (data not shown).

Cells were tested for reactivity to beef insulin in association with the appropriate class II MHC molecules ($A_b^b A_b^k$)¹⁶ (Fig. 2). Although the mutant-expressing cells did not generally release interleukin-2 (IL-2) before antigen stimulation, secretion was much enhanced after stimulation (Fig. 2). This was most notable at low concentrations of beef insulin ($40 \mu\text{g ml}^{-1}$), at which there was no response in control cells. As for parent cells, IL-2 production by the mutant cells was dependent on the antigen dose and required antigen presentation by the appropriate MHC molecules (Fig. 2a and inset; data not shown). The response conferred by the Lck mutant was qualitatively similar to that noted after transfection with the mouse CD4 gene^{10,15,16}

(comparison with a typical CD4-positive BI-141 clone is shown in Fig. 2b).

Protein phosphorylation on tyrosine residues is considered to be essential for T-cell activation^{17–19}. To understand better the mechanism by which the Lck mutant improved T-cell responsiveness, we examined the effects of the protein on TCR/CD3-mediated tyrosine phosphorylation (Fig. 3). Antiphosphotyrosine immunoblotting studies demonstrated that compared with control cells (lanes 2, 4 and 6), cells expressing the mutant showed a marked increase in tyrosine phosphorylation induced by antibody-mediated aggregation of TCR/CD3 (using anti-CD3 monoclonal antibody 145-2C11) (Fig. 3a, lanes 8, 10, 12 and 14).

The enhancement of the TCR/CD3-mediated tyrosine phosphorylation by the Lck mutant was also documented using the TCR-specific monoclonal antibodies F23.1 and H57-597 (Fig. 3b, compare lanes 8 and 9 with lanes 2 and 3). By contrast, no change in tyrosine protein phosphorylation was seen after antibody-mediated crosslinking of Thy-1.2 or CD45 on either control (lanes 4 and 5) or mutant cells (lanes 10 and 11) (Fig. 3c). The speed of tyrosine phosphorylation was similar in representative control and mutant-expressing cell lines (Fig. 3c), being maximal between 1 and 2 min (lanes 2, 3, 7 and 8) and decreasing by 5 min (lanes 4 and 9). Therefore the higher level of the TCR/CD3-induced tyrosine phosphorylation in the mutant-expressing cells probably reflects a higher stoichiometry of substrate phosphorylation and not altered kinetics.

We next compared the tyrosine phosphorylation induced by TCR-complex aggregation on the mutant-expressing cells with that resulting from TCR-CD4 coaggregation on a CD4-positive BI-141 transfectant. As reported by others¹⁹, coaggregation of TCR with CD4 (using F23.1-GK1.5 monoclonal antibody

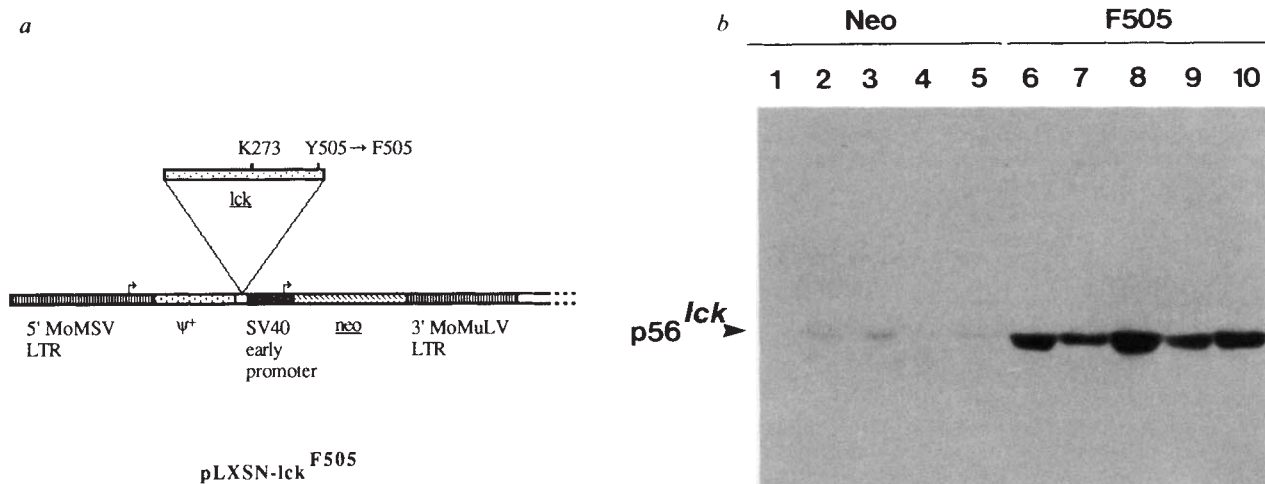


FIG. 1 Expression of F505 $p56^{lck}$ in BI-141 cells. **a**, Retroviral construct (pLXSN- lck^{F505}). The pLXSN retroviral vector²⁷ used to introduce lck complementary DNAs into BI-141 cells bears a convenient multiple cloning site flanked by a 5' Moloney murine sarcoma virus (MoMSV) long terminal repeat (LTR) and a 3' Moloney murine leukaemia virus (MoMuLV) LTR. The neomycin (*neo*) phosphotransferase gene (Tn5), which confers resistance to the aminoglycoside G418, is driven by the simian virus 40 (SV40) early promoter. The position of ψ^+ sequences is shown. Transcriptional orientations are indicated by arrows. K273 is a lysine residue at position 273 of $p56^{lck}$ and is part of the ATP-binding site. Y505 represents the major site of *in vivo* tyrosine phosphorylation of $p56^{lck}$, tyrosine 505. In our experiments, it has been replaced through site-directed mutagenesis by phenylalanine (F505)¹³. **b**, Lck immunoblot. Levels of $p56^{lck}$ were measured in clonal cell lines infected with retroviruses encoding either the neomycin phosphotransferase alone (lanes 1–5) or with F505 $p56^{lck}$ (lanes 6–10) using an Lck-specific immunoblot assay. The position of $p56^{lck}$ is indicated. Lanes: 1, Neo.1; 2, Neo.2; 3, Neo.3; 4, Neo.4; 5, Neo.5; 6, F505.3; 7, F505.7; 8, F505.9; 9, F505.10; 10, F505.12. Exposure, 6 h. The relative levels of $p56^{lck}$ overexpression in F505 cell lines (compared with the average of all Neo^r cell lines) are: F505.3,

6.5-fold; F505.7, 4.6-fold; F505.9, 11.1-fold; F505.10, 5.0-fold; F505.12, 9.0-fold.

METHODS. The F505 mutant lck cDNA has been described¹³. Retroviral infection of BI-141 cells was as before^{13,28}. Cells were selected for growth in media containing G418 ($750 \mu\text{g ml}^{-1}$) and monoclonal cell lines established by limiting dilution. For Lck immunoblotting, equivalent numbers of cells were directly lysed in sample buffer¹³. Lysates were resolved on 8% SDS-PAGE gels and transferred onto nitrocellulose membranes as described¹³. Immunoblotting was as described¹³ using either a polyclonal rabbit anti-Lck serum generated against a TrpE fusion protein bearing residues 5–148 of mouse $p56^{lck}$ (unpublished; construct²⁹ provided by R. Perlmutter, University of Washington) or a previously described rabbit antiserum generated against a synthetic peptide corresponding to amino acids 39–64 of the murine $p56^{lck}$ sequence¹. Both antisera are highly specific for $p56^{lck}$ and gave similar results (our unpublished data; data not shown). For quantitation, bands were cut from nitrocellulose and counted in a γ counter. The presence of equivalent amounts of proteins in each lane was confirmed by staining of nitrocellulose filters with amido black (data not shown).

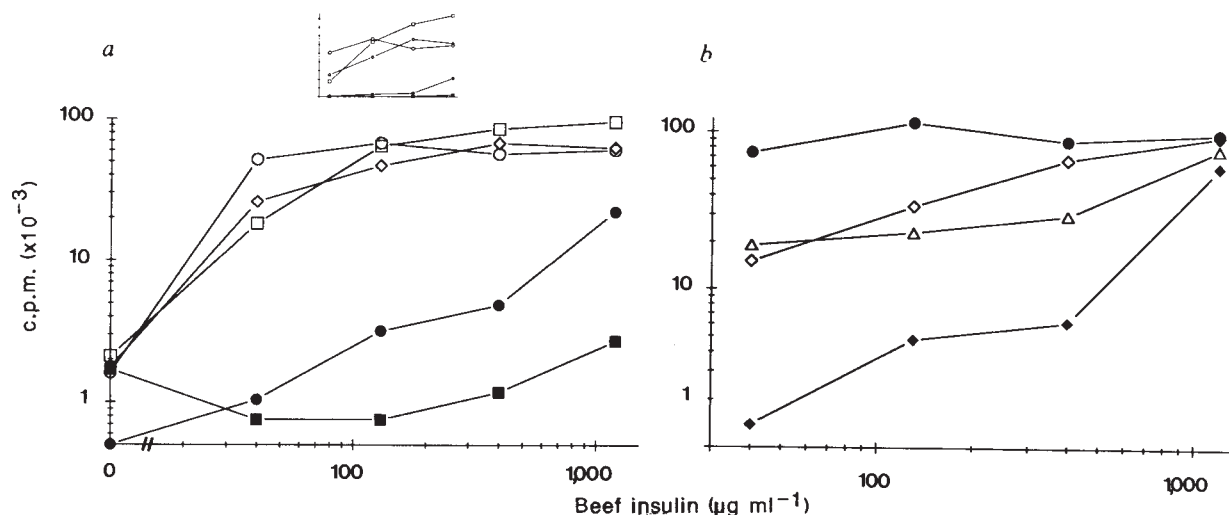
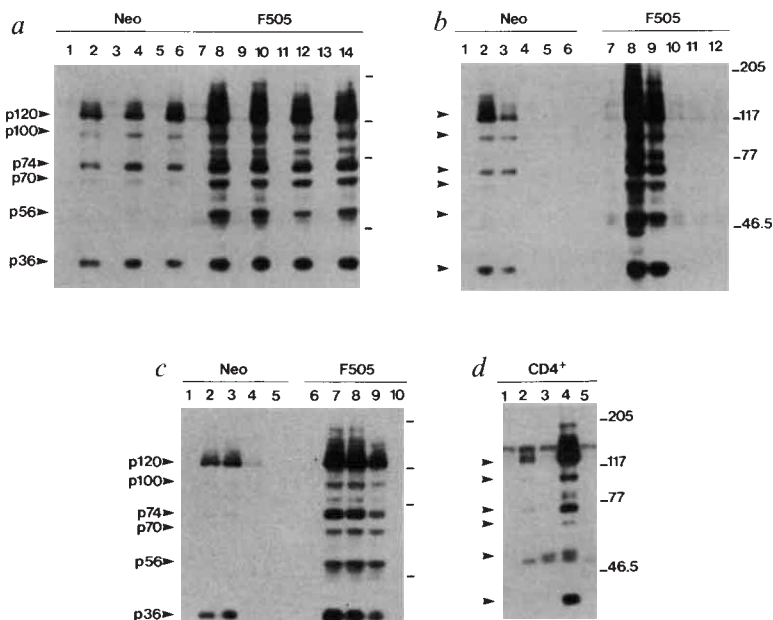


FIG. 2 Enhancement of antigen-induced IL-2 production by F505 p56^{lck}. IL-2 production in response to various concentrations of beef insulin in association with A₂A₂ class II MHC was tested in F505 p56^{lck}-expressing BL-141 cells and compared with that of other BL-141 derivatives. IL-2 assays were performed using the IL-2-sensitive cell line HT-2. Representative results of 13 experiments are shown and the enhancement of antigen-induced IL-2 production by F505 p56^{lck} expression has been highly reproducible throughout. Results are represented as counts per minute (c.p.m.) of ³H-thymidine incorporation. Ordinate, c.p.m. of [³H]thymidine incorporation (logarithmic scale); abscissa, concentration of beef insulin (μg ml⁻¹) (logarithmic scale). Inset, ordinate in linear scale. Similar results were obtained with two other randomly selected independent F505 p56^{lck} cell

lines (F505.1 and F505.10) (data not shown). *a*, Neo.1, ■; Neo.2, ●; F505.3, □; F505.7, ○; F505.9, ◇. Spontaneous IL-2 production is shown as the zero point on the abscissa. *b*, BL-141, ◆; F505.9, ◇; F505.12, △, CD4.C5, ●. METHODS. BL-141 cells (1 × 10⁵ per well) were plated in triplicate with 1 × 10⁵ MHC class II A₂A₂-transfected L cells (provided by R. Germain, NIH, Bethesda, Maryland) in 96-well flat-bottom tissue culture plates containing serial dilutions of beef insulin in a final volume of 200 μl tissue culture medium^{10,16}. After incubation at 37 °C for 24 h, supernatants were collected and assayed for IL-2 content. This was performed by measuring [³H]-thymidine incorporation in 1 × 10⁴ IL-2-dependent HT-2 indicator cells^{10,30}. Control cultures were without addition. Mouse CD4 (CD4.C5)-transfected BL-141 cells have been described¹⁰.

FIG. 3 Enhancement of TCR/CD3-induced tyrosine protein phosphorylation signal by F505 p56^{lck}. Anti-phosphotyrosine immunoblots. *a*, Effects of F505 p56^{lck} expression on TCR/CD3-induced tyrosine phosphorylation. Crosslinking was performed with anti-CD3 epsilon monoclonal antibody 145-2C11 and rabbit anti-hamster (RAH) IgG for 1 min at 37 °C. Lanes: 1 and 2, Neo.1; 3 and 4, Neo.2; 5 and 6, Neo.3; 7 and 8, F505.7; 9 and 10, F505.9; 11 and 12, F505.10; 13 and 14, F505.12. Lanes: 1, 3, 5, 7, 9, 11 and 13, RAH IgG alone; 2, 4, 6, 8, 10, 12 and 14, monoclonal antibody 145-2C11 with RAH IgG. The positions of relative molecular mass markers are indicated on the right and those of the major tyrosine phosphorylation substrates on the left. Quantitative analyses showed that the TCR/CD3-induced tyrosine phosphorylation of p120, p100 and p36 was enhanced two- to three-fold by F505 p56^{lck}; that of p74, two- to four-fold; p70, four- to fivefold; and p56, five- to ninefold (data not shown). Exposure: 6 h. *b*, Effects of crosslinking of other cell-surface epitopes. Crosslinking was performed for 1 min at 37 °C. Lanes: 1–6, Neo.1; 7–12, F505.9. Lanes: 1 and 7, untreated control; 2 and 8, anti-V_β8 antibody F23.1 plus RAM IgG; 3 and 9, anti-TCR antibody H57.597 plus RAH IgG; 4 and 10, anti-Thy-1.2 antibody G7; 5 and 11, anti-CD45 antibody M1/89.18.7 plus RAR IgG; 6 and 12, RAM IgG alone. Exposure, 6 h. *c*, Time-course experiments. Crosslinking was performed with anti-CD3 antibody 145-2C11 and RAH IgG for variable periods of time at 37 °C. Lanes: 1–5, Neo.1; 6–10, F505.9. Lanes: 1 and 6, untreated control; 2 and 7, 145-2C11 plus RAH IgG for 1 min; 3 and 8, 145-2C11 plus RAH IgG for 2 min; 4 and 9, 145-2C11 plus RAH IgG for 5 min; 5 and 10, RAH IgG for 5 min. Exposure 6 h. *d*, Effects of coaggregation of TCR-CD3 with CD4 on tyrosine protein phosphorylation. Antibody-mediated crosslinking of BL-141 cells transfected with CD4 was done for 1 min at 37 °C. Lanes: 1, untreated control; 2, anti-TCR antibody F23.1 plus rabbit anti-mouse (RAM) IgG; 3, anti-CD4 antibody GK1.5 plus rabbit anti-rat (RAR) IgG; 4, F23.1–GK1.5 heteroconjugates plus RAM IgG; 5, RAM IgG alone. Exposure, 6 h. Quantitative analyses of individual substrates revealed that TCR complex-induced tyrosine phosphorylation was enhanced between four- and sevenfold by coaggregation of TCR-CD3 with CD4. F23.1–GK1.5 heteroconjugates did not enhance the TCR-induced tyrosine phosphorylation



signal of CD4-negative BL-141 cells (data not shown).

METHODS. Antibody-mediated crosslinking and antiphosphotyrosine immunoblots were done as described previously^{5,6}, except that a polyclonal rabbit antiphosphotyrosine serum³¹ was used for these immunoblots (provided by T. Pawson, Mount Sinai Hospital Research Institute, Toronto). Bands were quantitated as described above. The presence of equivalent amounts of proteins in each lane was confirmed by amido-black staining of nitrocellulose filters (data not shown). The monoclonal antibodies have been previously reported^{32–37}. (F23.1–GK1.5 heteroconjugates were provided by T. Owens, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada.)

heteroconjugates) caused a marked increase in the TCR-mediated tyrosine phosphorylation (Fig. 3d, compare lane 4 with lane 2). This response has been ascribed to the activation and approximation of CD4-associated $p56^{lck}$ (ref. 19). Importantly, the substrates phosphorylated as a result of CD4 coaggregation had the same relative molecular masses as those phosphorylated by expression of the Lck mutant (Fig. 3a-c), although the amount of phosphorylated substrates resulting from CD4 coaggregation was generally greater (four- to sevenfold) than that resulting from expression of the mutant (two- to fivefold). The nature of these substrates remains undefined.

To determine whether the enhanced responsiveness of the mutant-expressing cells was directly attributable to the elevated tyrosine kinase activity of the mutant protein, we analysed the effects of expression of $p56^{lck}$ polypeptides having normal or abolished tyrosine protein kinase activity (Fig. 4). BI-141 cells were infected with retroviruses encoding either wild-type $p56^{lck}$ or a kinase-deficient F505 $p56^{lck}$ variant (R273F505 $p56^{lck}$; Fig. 4). Three wild-type and three R273F505 $p56^{lck}$ -expressing cell lines were analysed further. These cells respectively contained 5- to 14-fold (Fig. 4a, lanes 7-9) and 10- to 16-fold (lanes 10-12) higher amounts of $p56^{lck}$ than control cells (lanes 1-3). These levels of $p56^{lck}$ were comparable with those observed in F505 $p56^{lck}$ -expressing cells (lanes 4-6). All clones had unaltered

levels of TCR, CD3, Thy-1.2 and CD45 surface expression and remained CD4-negative (data not shown).

The responsiveness of these cells to beef insulin was compared with that of parental, control and F505 $p56^{lck}$ -expressing cells (Fig. 4b). None of the cells expressing R273F505 $p56^{lck}$ showed an enhanced antigen-induced IL-2 response. The wild-type $p56^{lck}$ -overexpressing cell lines also generally displayed an unaltered phenotype, except for wt21 which had a marginally improved IL-2 response. We also evaluated the effects of expression of these $p56^{lck}$ polypeptides on TCR/CD3-induced tyrosine phosphorylation. Antiphosphotyrosine immunoblot analyses (Fig. 4c) demonstrated that neither overexpression of wild-type $p56^{lck}$ (lanes 7-12) nor expression of R273F505 $p56^{lck}$ (lanes 13-18) greatly enhanced the TCR/CD3-associated tyrosine phosphorylation; wt 21 cells (lanes 11 and 12) showed slightly more TCR/CD3-induced tyrosine phosphorylation than control cells (lanes 1-4).

These data show that T-cell responsiveness is enhanced by expression of a constitutively activated form of the lymphocyte-specific tyrosine protein kinase $p56^{lck}$. This is indicated by an increase in both the antigen-induced IL-2 response and the tyrosine protein phosphorylation associated with the TCR complex. The results provide direct evidence that the interactions of CD4 and CD8 with $p56^{lck}$ are likely to be relevant to the functions of these molecules in T-cell activation. But CD4 and CD8 may also act as adhesion molecules, enhancing physical interactions between the T cell and the antigen-presenting cell²⁰ as well as coreceptors for the TCR-CD3 complex^{10,21-23}. This may explain why the enhancement of T-cell responsiveness by F505 $p56^{lck}$ is frequently less than that conferred by CD4 (this report; our unpublished data). Our results show that antigen receptor stimulation is required to observe the manifestations of F505 $p56^{lck}$ expression in BI-141 cells. This observation suggests a link between the antigen receptor signalling pathway and $p56^{lck}$. Our findings also indicate that activation of $p56^{lck}$ (for example, through CD45-mediated dephosphorylation of tyrosine 505; ref. 24) can directly enhance antigen receptor-mediated signals. Although the intersection of $p56^{lck}$ -derived signals with the TCR signalling pathway may be less efficient without the function of the CD4/CD8 coreceptor, such a process may be important for cells such as double-negative thymocytes, $\gamma\delta$ T-cells and natural killer cells^{6,25,26}, which express $p56^{lck}$ but lack expression of CD4 and CD8 (A.V., unpublished data).

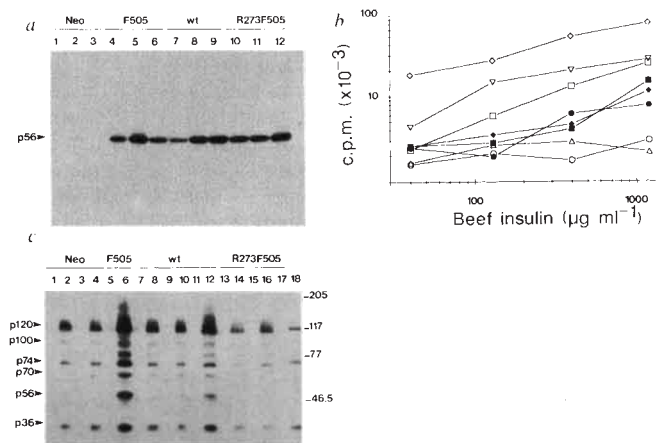


FIG. 4 Effects of wild-type (wt) and kinase-negative $p56^{lck}$ polypeptides on T-cell responsiveness. *a*, Lck protein immunoblot. As in Fig. 1b. Lanes: 1, Neo.1; 2, Neo.3; 3, Neo.4; 4, F505.7; 5, F505.9; 6, F505.10; 7, wt.8; 8, wt.10; 9, wt.21; 10, R273F505.3; 11, R273F505.17; 12, R273F505.19. Exposure, 8 h. The position of $p56^{lck}$ is indicated. Levels of $p56^{lck}$ expression were quantified as described above and (relative to the average of Neo cells) are as follows: wt.8, 5.3; wt.10, 14.2; wt.21, 13.4; R273F505.3, 12.8; R273F505.17, 10.5; R273F505.19, 16.6. *b*, Antigen responsiveness. As in Fig. 2. BI-141, ○; Neo.1, □; F505.9, △; wt.8, △; wt.21, ▽; R273F505.3, ●; R273F505.17, ◆; R273F505.19, ■. *c*, Effects of TCR-CD3 aggregation on tyrosine protein phosphorylation. As in Fig. 3a. Crosslinking was performed with monoclonal antibody 145-2C11 and RAH IgG for 1 min at 37°C. Lanes: 1 and 2, Neo.1; 3 and 4, Neo.2; 5 and 6, F505.9; 7 and 8, wt.8; 9 and 10, wt.10; 11 and 12, wt.21; 13 and 14, R273F505.3; 15 and 16, R273F505.17; 17 and 18, R273F505.19. Lanes: 1, 3, 5, 7, 9, 11, 13, 15 and 17, RAH IgG alone; 2, 4, 6, 8, 10, 12, 14, 16 and 18, 145-2C11 plus RAH IgG. Exposure, 3 h. Quantitative studies showed that the TCR/CD3-induced tyrosine phosphorylation of individual substrates did not significantly differ between control cells and cells overexpressing wild-type $p56^{lck}$ or expressing R273F505 $p56^{lck}$, except for wt 21 cells which demonstrated between a two- and threefold enhancement of tyrosine phosphorylation of these various substrates (data not shown).

METHODS. A kinase-deficient variant of F505 $p56^{lck}$ was engineered by replacing lysine 273 (within the ATP-binding site) by an arginine using oligonucleotide-directed mutagenesis (N.A. and A.V., unpublished data). We have previously shown that this mutant (R273F505 $p56^{lck}$) is devoid of tyrosine protein kinase activity and is incapable of oncogenic transformation of rodent fibroblasts (N.A. and A.V., unpublished observations). Generation of expression vectors, retrovirus stocks and infection of BI-141 cells were done as outlined in Fig. 1.

Received 3 December 1990; accepted 15 January 1991.

- Veilleille, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **55**, 301-308 (1988).
- Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5190-5194 (1988).
- Veilleille, A., Bookman, M. A., Horak, E. M., Samelson, L. E. & Bolen, J. B. *Nature* **338**, 257-259 (1989).
- Luo, K. & Sefton, B. M. *Molec. cell. Biol.* **10**, 5305-5313 (1990).
- Veilleille, A., Bolen, J. B. & Bookman, M. A. *Molec. cell. Biol.* **9**, 4441-4446 (1989).
- Veilleille, A., Zuniga-Pflucker, J. C., Bolen, J. B. & Kruisbeek, A. M. *J. exp. Med.* **170**, 1671-1680 (1989).
- Sleekman, B. P. et al. *J. Immunol.* **141**, 49-54 (1988).
- Zamoyska, R. et al. *Nature* **342**, 278-281 (1989).
- Letourneur, F. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2339-2343 (1990).
- Miceli, M. C., von Hoegen, P. & Parnes, J. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2623-2627 (1991).
- Amrein, K. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4247-4251 (1988).
- Marth, J. D. et al. *Molec. cell. Biol.* **8**, 540-550 (1988).
- Abraham, N. & Veilleille, A. *Molec. cell. Biol.* **10**, 5197-5206 (1990).
- Reske-Kunz, A. B. & Rude, E. *Eur. J. Immunol.* **15**, 1048-1054 (1985).
- Bailhausen, W. G. et al. *J. exp. Med.* **167**, 1493-1498 (1988).
- Parnes, J. R., von Hoegen, P., Miceli, M. C. & Zamoyska, R. *Cold Spring Harb. Symp. quant. Biol.* **54**, 649-655 (1989).
- Mustelin, T., Coggeshall, K. M., Isakov, N. & Altman, A. *Science* **247**, 1584-1587 (1990).
- June, C. H. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7722-7726 (1990).
- June, C. H., Fletcher, M. C., Ledbetter, J. A. & Samelson, L. E. *J. Immunol.* **144**, 1591-1599 (1990).
- Gay, D. et al. *Nature* **328**, 626-629 (1987).
- Anderson, P., Blue, M. L. & Schlossman, S. F. *J. Immunol.* **140**, 1732-1737 (1988).
- Gallagher, P. F., de St. Groth, B. F. & Miller, J. F. A. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10044-10048 (1989).
- Saizawa, K., Rojo, J. & Janeway, C. A. *Nature* **328**, 260-263 (1987).
- Ostergaard, H. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8959-8963 (1989).
- Reynolds, P. J. et al. *Molec. cell. Biol.* **10**, 4266-4270 (1990).
- Einspahr, K. J., Abraham, R. T., Dick, C. J. & Leibson, P. J. *J. Immunol.* **145**, 1490-1497 (1990).
- Miller, A. D. & Rosman, G. J. *BioTechniques* **7**, 980-990 (1989).
- Albritton, L. M., Tseng, L., Scadden, D. & Cunningham, J. M. *Cell* **57**, 659-666 (1989).
- Louie, R. R. et al. *J. Virol.* **62**, 4673-4679 (1988).
- Mossman, T. R. *J. Immunol. Meth.* **65**, 55-63 (1983).

31. Koch, C. A., Moran, M., Sadowski, I. & Pawson, T. *Molec. cell. Biol.* **9**, 4131–4140 (1989).
32. Staerz, U. D., Rammensee, H., Benedetto, J. D. & Bevan, M. J. *J. Immun.* **134**, 3994–4000 (1985).
33. Wilde, D. B., Marrack, P., Kappler, J., Dialynas, D. P. & Fitch, F. W. *J. Immun.* **131**, 2178–2183 (1983).
34. Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374–1378 (1987).
35. Kubo, R. T., Born, W., Kappler, J. W., Marrack, P. & Pigeon, M. *J. Immun.* **142**, 2736–2742 (1989).
36. Gunter, K. C., Malek, T. R. & Shevach, E. M. *J. exp. Med.* **159**, 716–730 (1984).
37. Springer, T., Galfre, G., Secher, D. S. & Milstein, C. *Eur. J. Immun.* **8**, 539–551 (1978).

ACKNOWLEDGEMENTS. We thank T. Owens, M. Julius, R. Germain, T. Pawson, R. Perlmutter, D. Miller, J. Bolen, R. Kubo and N. Sonenberg for reagents, and D. Davidson and L. Caron for reading the manuscript. This work was supported by the National Cancer Institute of Canada, the Medical Research Council of Canada and the Cancer Research Society Inc. (A.V.) and US NIH (J.R.P.). N.A. is supported by the Cancer Research Society Inc. M.C.M. is supported by a postdoctoral fellowship from the Arthritis Foundation. J.R.P. is an Established Investigator of the American Heart Association and A.V. is a Scholar of the Medical Research Council of Canada.

A discrete sequence in a platelet integrin is involved in ligand recognition

Stanley E. D'Souza*, Mark H. Ginsberg*, Gary R. Matsueda† & Edward F. Plow*

* Committee on Vascular Biology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

† Squibb Laboratories/Biology, Princeton University, Princeton, New Jersey 08544, USA

PLATELET membrane glycoprotein IIb-IIIa (gpIIb-IIIa; α IIb- β 3), the most prominent member of the integrin family of adhesion receptors on these cells, mediates platelet aggregation by binding fibrinogen and is critical in thrombosis and haemostasis^{1–5}. A short amino-acid sequence at the carboxy terminus of the γ chain of fibrinogen is recognized by gpIIb-IIIa (ref. 6) and peptides containing this sequence are selectively crosslinked to residues 294–314 of gpIIb (ref. 7). Here we show that an 11-residue peptide from this region of gpIIb inhibits platelet aggregation and binding of fibrinogen to platelets and to purified gpIIb-IIIa, and that it interacts directly with fibrinogen. These results implicate this segment of gpIIb-IIIa in the ligand-binding function of the receptor. Moreover, as this region is highly conserved among integrins, it may have a general function in ligand recognition by this broadly distributed family of adhesion receptors.

The glycoprotein gpIIb-IIIa is comprised of 1,770 amino acids^{8,9}, but we have focused on the peptide corresponding to residues 296–306 of gpIIb (TDVNGDGRHDL in one-letter amino-acid code; designated B12), because the fibrinogen γ -chain peptide crosslinks to this region and because it contains oxygenated residues that are spaced to conform to a calcium-binding motif^{9,10}. To investigate the function of B12, we first assessed its effect on platelet aggregation. Although thirty other peptides of 11–19 amino acids, representing sequences throughout gpIIb-IIIa, were inactive, B12 inhibited ADP-induced aggregation of washed human platelets in a dose-dependent manner (Fig. 1a). B12 concentrations as low as 5 μ M were inhibitory, and B12 abolished aggregation of twelve separate platelet preparations at concentrations ≥ 1 mM. Even with strong agonists such as thrombin (60 milliunits per ml) and collagen (5 μ g ml⁻¹), 1 mM B12 could partially inhibit (~40%) aggregation.

As shown in Fig. 1b, this inhibition of platelet aggregation can be ascribed to the effect of B12 on fibrinogen binding to the cells; the extent of inhibition depends on peptide concentration (50% at 0.4–0.5 mM; 4 experiments). Two different preparations of synthetic peptide and peptide purified further by HPLC behaved similarly. The effect was not due to chelation of calcium by the peptide as it persisted when the calcium in the assay was increased by 4-fold. Not only did 20 other gpIIb and gpIIIa peptides have little to no effect on ¹²⁵I-labelled fibrinogen binding to platelets at 1 mM, but a single conservative amino-acid

substitution (Glu for one of the Asp residues) resulted in a peptide (B12.1) that was much less potent (Fig. 1b). Moreover, B12 directly inhibited fibrinogen binding to purified gpIIb-IIIa. When purified gpIIb-IIIa was immobilized onto microtitre wells (see Fig. 3a legend), its specific binding to ¹²⁵I-labelled fibrinogen was inhibited by 58% by 0.5 mM B12, whereas four control gpIIb-IIIa peptides gave $\leq 9\%$ inhibition in the same experiment. Fibrinogen occasionally precipitated at high concentrations of B12, making construction of dose-response curves difficult. But at 0.2 mM B12 there was no precipitation, and binding inhibition of ¹²⁵I-labelled fibrinogen to immobilized gpIIb-IIIa was $64 \pm 18\%$, with a range of 47 to 89% in four experiments.

If B12 constitutes an element of the ligand-recognition site in gpIIb-IIIa, then we should be able to demonstrate its direct interaction with fibrinogen. Indeed, when we immobilized B12 onto the wells of microtitre plates, ¹²⁵I-labelled fibrinogen bound to this peptide 10–12 times more effectively than it did to other gpIIb-IIIa peptides (Fig. 2a). Fibrinogen also bound to B12.1 (E $\rightarrow$ D), but not as well: in four experiments, B12.1 bound 68% less fibrinogen than did B12. As shown in Fig. 2b, nonlabelled fibrinogen produced a concentration-dependent inhibition of ¹²⁵I-labelled fibrinogen binding to B12, which is evidence for a saturable interaction of the protein with the peptide; 50% inhibition of specific binding was observed at $0.3 \pm 0.08 \mu$ M (3 experiments).

The specificity of the interaction of B12 with ¹²⁵I-labelled fibrinogen is evident from Table 1: at a concentration of unlabelled fibrinogen that inhibited ¹²⁵I-fibrinogen binding by $>80\%$, the unrelated proteins albumin or thyroglobulin gave minimal inhibition. ¹²⁵I-fibrinogen binding to B12 is dependent on divalent ions, because 5 mM EDTA abolished the interaction. Binding was comparable in the presence of either Ca²⁺ or Mg²⁺: 16,600 and 15,900 c.p.m. ¹²⁵I-fibrinogen bound in the presence of 1 mM Ca²⁺ or Mg²⁺ respectively (2 experiments). Two γ -chain peptides inhibited the interaction of fibrinogen with immobilized B12, so also did a monoclonal antibody (4A5)

TABLE 1 Specificity of the interaction of ¹²⁵I-labelled fibrinogen with immobilized B12 peptide

Inhibitor	Concentration (μ M)	% Inhibition of ¹²⁵ I-fibrinogen binding to B12*
None		0
Fibrinogen	2	82 ± 1.2
Albumin	2	12 ± 4
Thyroglobulin	2	14 ± 4
EDTA	5,000	90 ± 8
HHLGGAKQAGDV†	200	57 ± 11
KYGGHHLGGAKQAGDV‡	200	54 ± 12
RGDS§	200	52 ± 15
GRGES§	200	16 ± 7
YAVTGRGDSPASSK	200	61 ± 16
4A5 antibody¶	10	75 ± 8
Control antibody¶	10	15 ± 4

* The binding of ¹²⁵I-labelled fibrinogen to the B12 peptide immobilized on microtitre plates was measured as detailed in the legend to Fig. 2.

† A representative γ -chain peptide corresponding to the sequence at the carboxy terminus of the γ -chain of human fibrinogen. This peptide inhibits fibrinogen binding to platelets⁶.

‡ The γ -chain peptide that was shown to be specifically crosslinked to gpIIb 294–314 by a heterobifunctional reagent⁷.

§ RGDS inhibits fibrinogen binding to platelets^{12,15}, whereas the conservative substitution in GRGES results in a much less potent inhibitor¹².

|| A peptide corresponding to a sequence in human fibronectin and containing an RGD sequence.

¶ 4A5 is a monoclonal antibody reactive with the carboxy terminus of the γ chain of fibrinogen¹¹; the control antibody reacts with a different region of fibrinogen.

raised against the carboxy-terminal region of the γ chain of fibrinogen¹¹. Peptides containing the sequence Arg-Gly-Asp (RGD in one-letter code) prevent fibrinogen and other adhesive proteins from binding to platelets and purified gpIIb-IIIa (refs 12-16); we found that two RGD-containing peptides could also inhibit the B12-fibrinogen interaction, whereas GRGESP was less effective. Specific fibrinogen binding (that component of ¹²⁵I-fibrinogen binding inhibitable by excess nonlabelled fibrinogen) to B12 inhibited by RGD peptides at 0.2 mM was $75.2\% \pm 20.1$, with a range of 55.4-100% in 6 experiments; for γ -chain peptide (0.2 mM) it was $78.9\% \pm 20.6$, with a range of

52.1-100% (6 experiments). These observations are consistent with the hypothesis that RGD and γ -chain peptides interact with the same or mutually exclusive sites in gpIIb-IIIa (refs 17, 18).

We next investigated the effects on ligand binding of an antibody raised against the B12 region to assess its role in receptor function. The antibody was obtained by passing sera of rabbits immunized with gpIIb-IIIa over a B12 affinity column and then eluting the bound antibody with acid; this anti-B12 reacted with the peptide on microtitre wells at a 1/1,000 dilution and not with other gpIIb-IIIa peptides at a 1/10 dilution. It reacted exclusively with gpIIb when used for immunoblotting platelet extracts but did not after preincubation of the antibody with B12, although it did after preincubation with a control gpIIb peptide. Anti-B12 blocked fibrinogen binding to the peptide and to purified gpIIb-IIIa (Fig. 3a). In addition, anti-B12 also inhibited ¹²⁵I-labelled fibrinogen binding to intact platelets (Fig. 3b). At a concentration of anti-B12 that inhibited fibrinogen binding to platelets by >80%, three control anti-peptide antibodies had no effect.

In sum, our data implicate residues 296-306 of gpIIb in the ligand-binding function of gpIIb-IIIa. This 11-amino-acid peptide not only inhibits receptor function but also binds directly to the fibrinogen ligand. This interaction mimicks ligand binding to purified gpIIb-IIIa and to platelets with respect to divalent

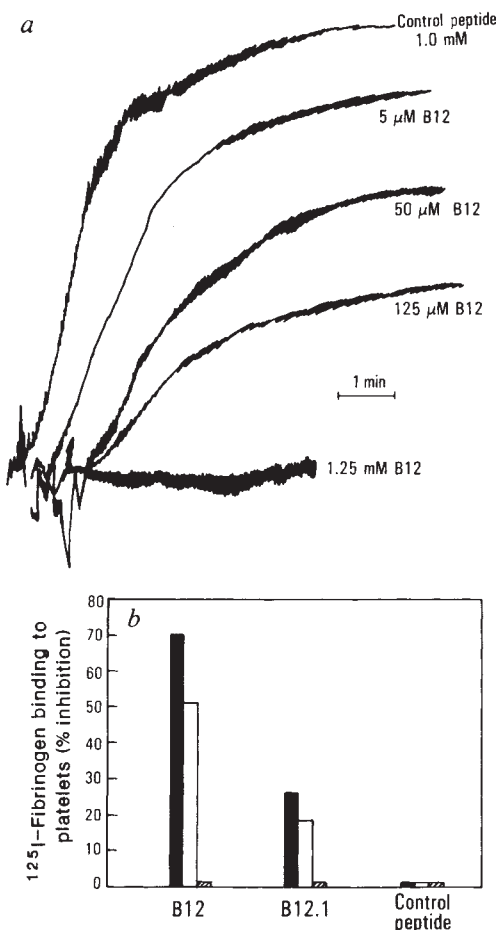


FIG. 1 a, Effect of the B12 peptide, gpIIb 296-306, on platelet aggregation. The indicated concentrations (final) of B12 were added to washed human platelets (10^8 per ml) in Tyrode's buffer (pH 7.4) (refs 12, 27, 28) containing 1 mM Ca^{2+} and 300 nM fibrinogen. Aggregation was initiated with 10 μ M ADP. The control peptide shown is residues 461-471 of GP1Ib and is representative of 30 gpIIb-IIIa peptides tested. At the highest concentration of B12, no platelet aggregation was detected even after 10 min. b, Effect of B12 on fibrinogen binding to platelets. ¹²⁵I-labelled fibrinogen (60 nM) was added to washed platelets stimulated with ADP (10 μ M) and binding was measured at 22 °C in the presence or absence of B12 (TDVNGDGRHDL), B12.1 (TDVNGEGRHDL) and control peptides at final concentrations of 1.0 mM (solid), 0.5 mM (open) and 0.05 mM (hatched). The bound radiolabelled fibrinogen was quantitated and expressed as a per cent of fibrinogen bound in the absence of added peptide. The control peptides shown were: gpIIb 24-33, 363-374, 430-439 and 530-544. The inhibition by B12 was time-dependent; results shown were obtained when the ¹²⁵I-labelled fibrinogen was preincubated with B12 for 30 min before addition of the ADP-stimulated platelets, and binding of the radiolabelled ligand was measured after 10 min. Although preincubation with B12 was not necessary for inhibition, as the incubation of ¹²⁵I-labelled fibrinogen with the platelets was extended, the inhibitory activity of B12 diminished. This decrease in B12 activity with time might arise from competition between B12 and the platelet receptor for the same site on fibrinogen, with the higher-affinity platelet receptor ultimately displacing the peptide from the ligand, particularly as the binding of fibrinogen to the platelet becomes irreversible³.

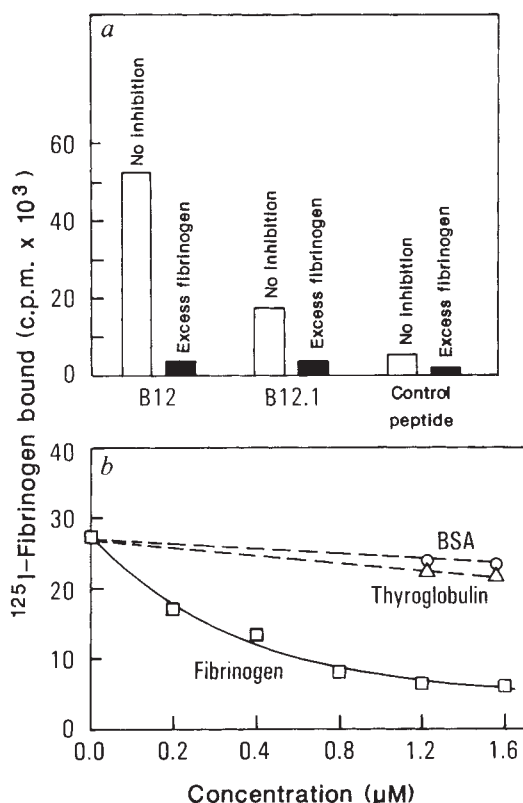


FIG. 2 a, Fibrinogen binding to immobilized B12 peptide, B12.1 or control gpIIb-IIIa peptides. The control peptide was gpIIb 530-544 and was representative of 8 other gpIIb and gpIIIa peptides (100 μ l at 1 mM) added to 96-well Immulon-2 microtitre plates (Dynatech) for 24 h at 4 °C. The peptides were removed and the plates washed with HEPES-Tyrode's buffer, pH 7.4, containing 0.05% Tween-20. Plates were post-coated with 3% BSA for 1 h at 37 °C and then washed again. ¹²⁵I-labelled fibrinogen (50 μ l at 50 nM) was allowed to bind to the peptide-coated microtitre wells for 24 h at 22 °C; plates were then extensively washed and counted for bound radioactivity. Labelled fibrinogen was displaced by unlabelled fibrinogen (2.5 μ M), which was indicative of saturable binding. b, Specificity of fibrinogen binding to the B12 peptide. Analysis was as in panel a. The unlabelled fibrinogen, BSA or thyroglobulin was added at the indicated concentration simultaneously with the ¹²⁵I-labelled fibrinogen (50 nM).

ion dependence and inhibition by RGD and γ -chain peptides. Previous crosslinking studies with RGD peptides implicated gpIIb residues 109–171 in ligand recognition¹⁹, and a single point mutation within this gpIIb region abolished ligand binding²⁰. Also, gpIIb 210–221 has also been implicated in ligand binding²¹. Thus at least two contact sites other than gpIIb 296–306 might exist between the receptor and the ligand. Each of the three possible regions contains five of the seven amino-acid residues appropriately spaced to form a calcium-binding motif such as that found in troponin and calmodulin^{10,22}. Coordination of a divalent ion between these amino acids and the ligand may be an essential element in binding. As several integrin receptors with the same β subunit recognize different ligands^{1,23}, the α subunit must be important in imparting ligand specificity²⁴. Nevertheless, residues 296–306 of gpIIb are highly conserved within the integrins. Whether the individual differences in the amino-acid sequences of this specific region are sufficient to account for the diversity in ligand recognition of integrins, or whether other regions of these receptors also influence ligand specificity is unknown. It is notable that this region of gpIIb is unique among the α subunits in containing a DGR sequence, the inverse of RGD. Peptides with this inverse sequence do inhibit ligand binding to certain integrins^{25,26} and could contribute to the ligand binding functions of the 296–306

sequence of gpIIb. We are now in a position to compare the interaction of gpIIb–IIIa ligands such as fibrinogen, fibronectin and von Willebrand factor with B12 and to assess the ligand-binding functions of the homologous peptide sequences of the other integrins. □

Received 27 August; accepted 27 December 1991.

- Hynes, R. O. *Cell* **48**, 549–554 (1987).
- Phillips, D. R., Charo, I. F., Parise, L. V. & Fitzgerald, L. A. *Blood* **71**, 831–843 (1988).
- Plow, E. F. & Ginsberg, M. H. in *Progress in Hemostasis and Thrombosis* Vol. 9 (ed. Collier, B.S.) 117–156 (Saunders, Philadelphia, 1989).
- Yasuda, T. et al. *J. clin. Invest.* **81**, 1284–1291 (1988).
- Akiyama, S. K., Nagata, K. & Yamada, K. M. *Biochim. biophys. Acta* **1031**, 91–110 (1990).
- Kloczewska, M., Timmons, S., Lukas, T. J. & Hawiger, J. *Biochemistry* **23**, 1767–1774 (1984).
- D'Souza, S. E., Ginsberg, M. H., Burke, T. A. & Plow, E. F. *J. biol. Chem.* **265**, 3440–3446 (1990).
- Fitzgerald, L. A., Steiner, B., Rall, S. C. Jr, Lo, S.-S. & Phillips, D. R. *J. biol. Chem.* **262**, 3936–3939 (1987).
- Poncz, M. et al. *J. biol. Chem.* **262**, 8476–8482 (1987).
- Kretsinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248**, 3313–3326 (1973).
- Matsueda, G. R., Ball, E. L. & Bernatowicz, M. S. *FASEB J.* **2**, 6480 (1988).
- Plow, E. F., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. A. & Ginsberg, M. H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8057–8061 (1985).
- Ginsberg, M. H., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. & Plow, E. *J. biol. Chem.* **260**, 3931–3936 (1985).
- Haverstick, D. M., Cowan, J. F., Yamada, K. M. & Santoro, S. A. *Blood* **66**, 946–952 (1985).
- Gartner, T. K. & Bennett, J. S. *J. biol. Chem.* **260**, 11891–11894 (1985).
- Parise, L. V. & Phillips, D. R. *J. biol. Chem.* **260**, 10698–10707 (1985).
- Lam, C.-T. et al. *J. biol. Chem.* **262**, 947–950 (1987).
- Bennett, J. S., Shattil, S. J., Power, J. W. & Gartner, T. K. *J. biol. Chem.* **263**, 12948–12953 (1988).
- D'Souza, S. E., Ginsberg, M. H., Burke, T. A., Lam, S. C.-T. & Plow, E. F. *Science* **242**, 91–93 (1988).
- Loftus, J. C. et al. *Science* **249**, 915–918 (1990).
- Charo, I. F., Nannizzi, L., Phillips, D. R., Hsu, M. A. & Scarborough, R. M. *Blood* **74**, 497a (1989).
- Kobayashi, T., Takagi, T., Konishi, K. & Wnuk, W. *J. biol. Chem.* **264**, 18247–18259 (1989).
- Hemler, M. E. *A. Rev. Immun.* **8**, 365–400 (1990).
- Ginsberg, M. H. et al. *J. biol. Chem.* **262**, 5437–5440 (1987).
- Yamada, K. M. & Kennedy, D. W. *J. cell. Physiol.* **130**, 21–28 (1987).
- Akiyama, S. K. & Yamada, K. M. *J. biol. Chem.* **260**, 10402–10405 (1985).
- Marguerie, G. A., Edgington, T. S. & Plow, E. F. *J. biol. Chem.* **255**, 154 (1980).
- Plow, E. F., Pierschbacher, M. D., Ruoslahti, E., Marguerie, G. & Ginsberg, M. H. *Blood* **70**, 110–115 (1987).
- Pytela, R., Pierschbacher, M. D., Ginsberg, M. H., Plow, E. F. & Ruoslahti, E. *Science* **231**, 1559–1562 (1986).

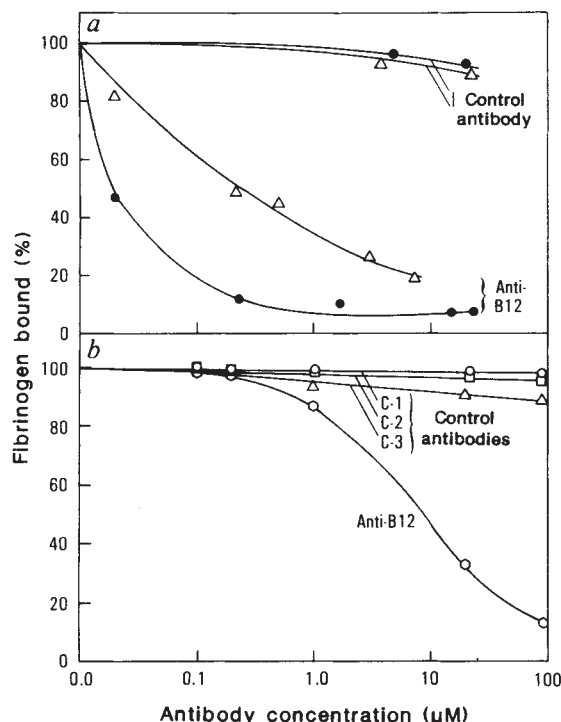


FIG. 3 *a*, Effect of antibodies to the B12 region of gpIIb–IIIa on fibrinogen binding to the B12 peptide and to purified gpIIb–IIIa. Rabbit antiserum raised against isolated B12 was passed over a B12–Sepharose column (2 mg peptide per ml Sepharose 4B), and bound antibody was eluted at pH 2.5. The control antibody was against gpIIb 1–13 and was immunopurified on a peptide column following the same protocol as for anti-B12. Antibodies at the indicated concentrations and ¹²⁵I-labelled fibrinogen (50 nM) were incubated with B12 (●) or gpIIb–IIIa (Δ), immobilized onto the wells of microtitre plates for 24 h at 22°. The B12 plates were prepared as described for Fig. 2, and the gpIIb–IIIa plates were prepared by coating the wells with 10 μg ml⁻¹ of gpIIb–IIIa, purified by affinity chromatography on KYGRGDS as described²⁹. Per cent binding was calculated relative to samples without antibodies. *b*, Effect of the antibodies on fibrinogen binding to platelets. ADP-stimulated, washed human platelets (10⁸ per ml), ¹²⁵I-labelled fibrinogen (60 nM) and the indicated concentration of antibodies were incubated for 30 min at 22°, and ligand binding measured as described^{12,27}. Antibodies against other peptides or regions of gpIIb–IIIa were used as controls.

Cell-surface regulation of β_1 -integrin activity on developing retinal neurons

K. M. Neugebauer* & L. F. Reichardt

Howard Hughes Medical Institute and Department of Physiology, University of California, San Francisco, California 94143-0724, USA

INTEGRINS are a family of $\alpha\beta$ heterodimeric receptors that mediate cell–cell and cell–substratum interactions. Integrin binding to extracellular ligands regulates cell adhesion, shape, motility, intracellular signalling and gene expression^{1–3}. Mechanisms that regulate integrin function are, therefore, central to the participation of integrins in a diverse set of cellular events. Here we report the identification of TASC, a monoclonal antibody to a novel epitope on the integrin β_1 subunit, which inhibits cell adhesion to vitronectin but promotes adhesion to laminin and collagen types I and IV. We show that developing retinal neurons that have lost responsiveness to laminin⁴ regain the ability to bind laminin in the presence of TASC. Thus, β_1 -class integrins are likely to occupy multiple affinity states that can be modulated at the cell surface.

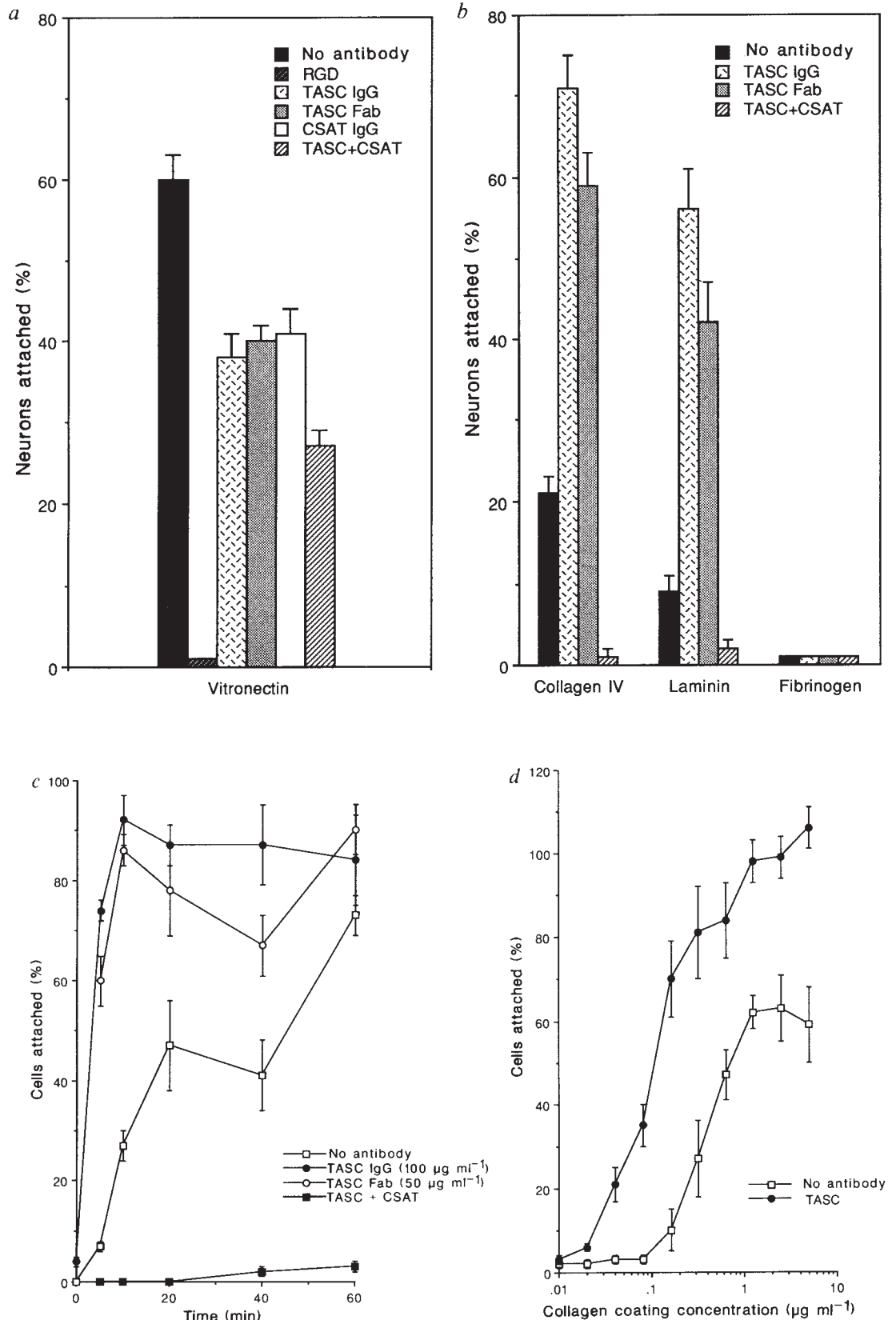
On developing neurons, β_1 -class integrins promote adhesion and growth cone motility in response to glycoprotein components of basal laminae and are, therefore, likely to play a prominent part in axon extension and pathfinding *in vivo*⁵. Retinal neurons from embryonic day six (E6) chicks express β_1 -class integrins that mediate adhesion and neurite outgrowth *in vitro* on purified laminin, collagen IV and fibronectin⁴. Embryonic day seven (E7) retinal neurons also attach to vitronectin

* Present address: Fred Hutchinson Cancer Research Center, M723 Basic Sciences, 1124 Columbia Street, Seattle, Washington 98104, USA.

FIG. 1 TASC inhibits E7 retinal neuron attachment to vitronectin and promotes attachment to laminin and collagens. *a*, Attachment to vitronectin is inhibited by $100 \mu\text{g ml}^{-1}$ GRGDSP and $200 \mu\text{g ml}^{-1}$ TASC, $50 \mu\text{g ml}^{-1}$ TASC Fab, and $50 \mu\text{g ml}^{-1}$ CSAT. *b*, Histograms of retinal neuron attachment to collagen IV (10 min), laminin (20 min) and fibrinogen (1 h). All antibodies are used at $50 \mu\text{g ml}^{-1}$. *c*, Time-course of E7 retinal neuron attachment to collagen I in the presence of no antibody, TASC IgG, TASC Fab and TASC IgG plus CSAT. *d*, Attachment of E7 retinal neurons to a collagen-I concentration curve at 40 min in the presence of no antibody or TASC at $25 \mu\text{g ml}^{-1}$. All values are expressed as a percentage of the attachment observed on poly-D-lysine which is considered to represent a signal of 100%.

METHODS. Vitronectin was purified from fetal bovine serum by chromatography on Heparin Sepharose Cl-6b (Pharmacia)²². Collagen I (Collaborative Research) was diluted to $10 \mu\text{g ml}^{-1}$ in 0.1 N acetic acid. Collagen IV (Collaborative Research) was diluted to $10 \mu\text{g ml}^{-1}$ in CMF-PBS. Laminin (diluted to $20 \mu\text{g ml}^{-1}$ in CMF-PBS) was purified from murine EHS tumour by I. de Curtis and M. J. Ignatius according to published procedures²³. Fibrinogen was the gift of Z. Ruggeri (Research Institute of the Scripps Clinic, La Jolla, California) and was used at $20 \mu\text{g ml}^{-1}$ in $\text{Ca}^{2+}/\text{Mg}^{2+}$ -containing PBS. Poly-D-lysine was from Sigma. IgGs were purified from ascites fluid on protein A/Sepharose Cl-4b (Pharmacia), and GRGDSP was purchased from Collaborative Research. TASC Fab fragments were generated from purified IgG by papain (Sigma) digestion followed by ion-exchange chromatography on DE-52 (Whatman). Cell attachment was quantitated as previously described⁴, except that adherent cells were stained with crystal violet and adsorption at 540 nm was subsequently measured²⁴. In each experiment, attachment to BSA-coated wells was subtracted, and all determinations were done in triplicate. To isolate TASC, integrin-rich chick

brain glial cells were used as the immunogen for monoclonal antibody production in mice²⁵. Clonal supernatants were screened for the ability to inhibit retinal neuron attachment to vitronectin.



(Fig. 1a). Attachment to vitronectin was abolished by an Arg-Gly-Asp-containing peptide indicative of integrin function⁶. The function-blocking monoclonal antibody CSAT to the integrin β_1 subunit⁷ inhibited attachment by 32%, implicating β_1 -class integrins in neuronal attachment to vitronectin (Fig. 1a). The TASC monoclonal antibody was isolated in a screen which used this attachment assay to select clones of interest. TASC IgG inhibited retinal neuron attachment to vitronectin by 37%, and Fab fragments of TASC IgG were also effective (Fig. 1a). The combined effects of TASC plus CSAT were stronger but did not completely block attachment to vitronectin, suggesting that additional vitronectin-binding integrins are expressed by these neurons.

TASC was subsequently found to promote attachment to other β_1 -class integrin ligands, including laminin, collagen I and collagen IV (Fig. 1b-d). E7 retinal neuron attachment to laminin was increased about fivefold and to collagen IV about threefold by both bivalent and monovalent TASC in short-term assays (Fig. 1b). The ability of TASC to promote attachment to collagen I was examined in detail. Both TASC IgG and Fab fragments accelerated E7 retinal neuron attachment to collagen I (Fig. 1c): whereas about 70% of the neurons adhered to collagen I in 1 h in the absence of TASC IgG, the same level of adhesion was achieved in the presence of TASC in 5 min. Because monovalent TASC increased attachment to all three ligands (Fig. 1b and c), its effects were not due to bivalent antibody-induced crosslinking of the TASC antigen. All of TASC's effects were β_1 -class integrin-specific, as the enhanced binding to laminin, collagen I and collagen IV was abolished by the function-blocking CSAT antibody (Fig. 1b and c). TASC also lowered the dose-dependence of attachment to collagen I by about fourfold (Fig. 1d). Note that TASC did not promote attachment to very low concentrations of ligand or to the BSA-blocked substrate, excluding the possibility that TASC itself had been adsorbed to the substrate. Furthermore, E7 retinal neurons did not attach to

TABLE 1 Antibody competition of adhesion to antibody substrates

Substrate antibody	Control attachment (%) Competing antibody		
	CSAT	JG22	TASC
CSAT	7 ± 2	7 ± 2	120 ± 18
JG22	5 ± 2	10 ± 2	131 ± 13
TASC	96 ± 11	90 ± 11	5 ± 1

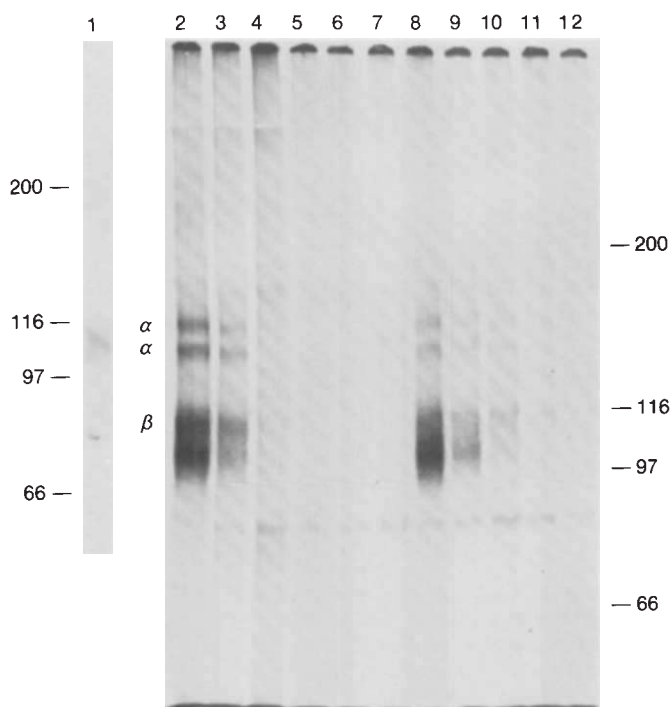
E7 retinal neuron attachment to substrates coated with 5 $\mu\text{g ml}^{-1}$ CSAT, TASC, and JG22 antibodies (see Fig. 1 for methods) was measured. Cells were incubated with the competing antibody for 2 min at room temperature, centrifuged onto the antibody substrates, incubated for 5 min and washed. The results (mean $\pm$ s.e.m.) are normalized to the attachment observed to each antibody in the absence of competing antibody. The s.e.m. for control attachment to CSAT was 14%, to JG22 8%, and to TASC 11%.

fibrinogen in the absence or presence of TASC (Fig 1b) even though β_1 -dependent attachment to fibrinogen has been detected in a chicken myeloblast cell line and is promoted by TASC (K.M.N. and L.F.R., manuscript in preparation). Thus, TASC does not seem to promote attachment by transforming the ligand specificity of β_1 -class integrin heterodimers expressed by a particular cell. It is also unlikely that TASC promoted the expression of new receptors on the cell surface, because its effects were detectable within minutes. Rather, TASC seems to increase the efficacy of ligand-specific binding by integrins that are already expressed.

Biochemical characterization of the TASC antigen revealed that TASC recognizes a novel epitope on the integrin β_1 subunit. In immunoblots of retinal cell extracts, TASC bound a band of relative molecular mass 105,000–110,000 (M_r 105–110K) (Fig. 2). From ³⁵S-labelled extracts of E7 retinal neurons, TASC IgG precipitated the same set of proteins as those precipitated by the anti- β_1 antibody CSAT (Fig. 2, lanes 2 and 8). These included

FIG. 2 TASC recognizes the integrin β_1 subunit. TASC binds a 105–110K band in an immunoblot of E7 retinal extracts (lane 1). Both TASC and CSAT immunoprecipitate the integrin β_1 subunit of 100K and 110K and associated α subunits at 135K and 145K. Immunodepletion analysis of the TASC antigen on E7 retinal neurons reveals that sequential precipitation of labelled extract with CSAT (lanes 2–5) removes the TASC antigen (lane 6). Overexposure of the autoradiogram did not reveal the presence of even minor amounts of either antigen following depletion (not shown). Antigens were unlikely to have been degraded during the course of the experiment, because another cell-surface glycoprotein, the integrin α_v subunit, was readily detected by immunoprecipitation from an aliquot of the same depleted sample (lane 7). Depletion with TASC (lanes 8–11) removes the CSAT antigen (lane 12).

METHODS. E7 retinas were extracted for 15 min in an equal volume of ice-cold $\text{Ca}^{2+}/\text{Mg}^{2+}$ -containing PBS plus 1% Triton X-100 and 2 mM PMSF. Supernatant (200 μg) was run under nonreducing conditions on a 6% gel and transferred to nitrocellulose. The blots were incubated with 20 $\mu\text{g ml}^{-1}$ TASC overnight at 4 °C, followed by alkaline phosphatase-conjugated second antibody (Promega), used and developed according to manufacturer's instructions. For immunoprecipitation, E7 retinal neurons were metabolically labelled with 100 $\mu\text{Ci ml}^{-1}$ [³⁵S] methionine/cysteine (NEN) for 16 h. Cells were collected and extracted as above. The supernatant was divided into two 200 μl aliquots, and 20 μg TASC or CSAT was added to each. Samples were rocked overnight at 4 °C and then precleared twice with 50 μl of Sepharose CL-4b (Pharmacia). Antibody-antigen complexes were precipitated with protein A-Sepharose CL-4B and washed extensively with 0.05M Tris buffer pH 8.0, 150 mM NaCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 0.5% Triton X-100. To deplete the extract of antigen, supernatants were repeatedly precipitated with 20 μg of each antibody. An additional protein A-Sepharose adsorption step was performed before adding the second antibody to ensure that all of the depleting antibody had been removed. The resulting supernatant was divided in half and precipitated with 20 μg of the second antibody. The monoclonal antibody to α_v was isolated in this laboratory²⁵. Pellets were extracted in sample buffer and separated by nonreducing SDS-PAGE on 6% gels. Gels were fixed, stained, treated with En³Hance (NEN), dried, and exposed to Kodak X Omat R film. M_r standards were obtained from Biorad.



proteins comigrating with the mature (110K) and immature (100K) forms of the β_1 subunit¹ in addition to the noncovalently associated α subunits (M_r 145 and 135K) previously characterized on retinal neurons⁴. As immunodepletion of the extract with TASC removed the CSAT antigen and vice versa (Fig. 2), both antibodies seem to recognize the same set of β_1 -class integrin heterodimers.

Because TASC and CSAT had different effects on cell adhesion, it seemed likely that the two antibodies recognize distinct epitopes on the β_1 subunit. To test this possibility, retinal neuron attachment to antibody-coated substrates was measured in the absence and presence of competing antibodies. Although CSAT competed for attachment to CSAT and TASC competed for attachment to TASC, neither antibody inhibited attachment to the other (Table 1). The monoclonal antibody JG22 binds an epitope on β_1 which overlaps with CSAT's⁸. Antibody JG22 competed for cell binding to CSAT and vice versa, but did not compete for binding to TASC (Table 1). Thus, CSAT and TASC recognize nonoverlapping epitopes on the integrin β_1 subunit. The observations that these two anti- β_1 antibodies have complementary effects on attachment to laminin and collagens and that TASC inhibits attachment to vitronectin suggest that two distinct sites on the β_1 subunit are involved in forming ligand-binding domains. Consistent with this structural prediction, evidence for the presence of two distinct ligand-binding domains on $\alpha_{IIb}\beta_3$ and $\alpha_4\beta_1$ heterodimers has been reported⁹⁻¹¹.

Purified laminin is a potent promoter of neurite outgrowth by early embryonic chicken retinal neurons and both attachment and neurite outgrowth depend on β_1 -class integrins⁴. These responses diminish with increasing developmental age, such that E11 retinal neurons do not attach or extend neurites on purified laminin. But β_1 -integrins are detectable on the surfaces of these neurons⁴. In addition, CSAT inhibits E11 retinal neurite out-

growth on astrocyte monolayers where the cell adhesion molecules NCAM and N-cadherin are also active¹². These observations suggest that E11 neurons express β_1 -class integrins that function in situations where multiple neurite-promoting receptors are engaged. Consistent with this, the number of ¹²⁵I-labelled laminin binding sites on the majority of retinal neurons remains relatively constant between E6 and E11, although retinal ganglion neurons (~10% of all retinal neurons) exhibit a decrease in the number of laminin receptors¹³. To test the possibility that these unresponsive cells continue to express functional integrin laminin receptors, their attachment to laminin was measured in the presence of TASC. Within 10 min, E11 retinal neurons had begun to attach to laminin in the presence of TASC, and 80% of the cells were attached in 40 min (Fig. 3). Thus, E11 retinal neurons express integrin receptors for laminin and the observed developmental changes must be due, at least in part, to mechanisms which lead to the reversible inactivation of receptor populations on the cell surface. This finding raises the possibility that laminin receptors continue to function on E11 neurons *in vivo* and that their function may be dynamically regulated by physiological events, such as cell contact and synapse formation.

Because TASC recognizes β_1 -class integrins and promotes their function with respect to attachment kinetics and dose-dependence of ligand, and in cells where their function has been downregulated, we propose that TASC binding increases receptor affinity for laminin, collagen I and collagen IV by inducing a conformational change in specific β_1 -class integrin heterodimers. Several monoclonal antibodies against integrin α subunits that promote heterodimer function have recently been described, suggesting that antibody-induced changes influence the conformation of associated α and β subunits¹⁴⁻¹⁶. Precedence for the rapid regulation of integrin avidity without a corresponding change in levels of receptor expression has been established in circulating blood cells and is likely to depend on intracellular signalling mechanisms^{3,17-20}. On macrophage, for example, interferon- γ and phorbol esters induce cell binding to laminin through the $\alpha_6\beta_1$ heterodimer, and binding is coincident with the phosphorylation of α_6 and association of the receptor with the cytoskeleton²¹. The results presented here show that β_1 -class integrins can occupy multiple functional states on the neuronal cell surface. This suggests that receptor avidity is also modulated on non-circulating cells whose movements within tissues may require rapid and dynamic regulation. □

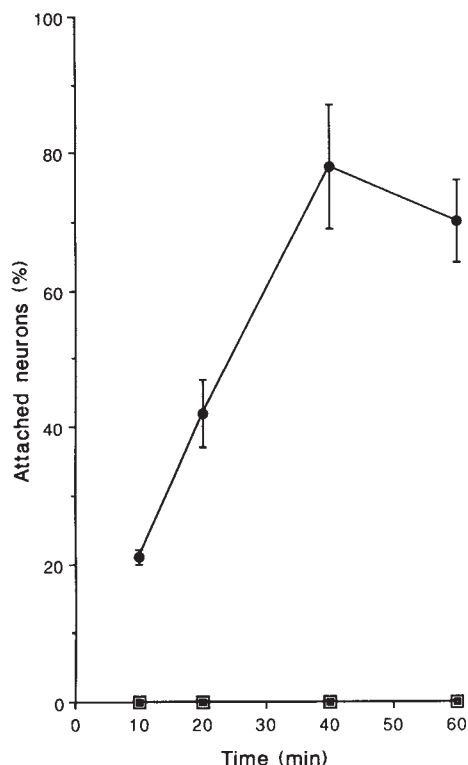


FIG. 3 E11 retinal neurons do not adhere to laminin in the absence of TASC at any time point up to 60 min (open squares). In the presence of 100 $\mu\text{g ml}^{-1}$ TASC (●), attachment to laminin is detectable within 10 min and peaks at 40 min with ~80% of the neurons attached compared to poly-D-lysine. TASC-induced attachment is completely blocked by the addition of CSAT (■).

Received 7 November; accepted 10 December 1990.

1. Akiyama, S. K., Nagata, K. & Yamada, K. M. *Biochim. biophys. Acta* **1031**, 91-110 (1990).
2. Hemler, M. E. *A. Rev. Immun.* **8**, 365-400 (1990).
3. Springer, T. A. *Nature* **346**, 425-434 (1990).
4. Hall, D. E., Neugebauer, K. M. & Reichardt, L. F. *J. Cell Biol.* **104**, 623-634 (1987).
5. Reichardt, L. F. & Tomaselli, K. J. *A. Rev. Neurosci.* **14**, 531-570 (1991).
6. Pierschbacher, M. D. & Ruoslahti, E. *Nature* **309**, 30-33 (1984).
7. Buck, C. A., Shea, E., Duggan, K. & Horwitz, A. F. *J. Cell Biol.* **103**, 2421-2428 (1986).
8. Horwitz, A. F. *et al.* in *Monoclonal Antibodies and Functional Cell Lines* (eds Kennett, R. H., Bechtol, K. B. & McKearn, T. J.) 103-118 (Plenum, 1984).
9. Santoro, S. A. & Lawing, W. J. *Jr Cell* **48**, 867-873 (1987).
10. Elices, M. J. *et al. Cell* **60**, 577-584 (1990).
11. Frelinger, A. L. *et al. J. Biol. Chem.* **265**, 6346-6352 (1990).
12. Neugebauer, K. M., Tomaselli, K. J., Lilien, J. & Reichardt, L. F. *J. Cell Biol.* **107**, 1177-1187 (1988).
13. Cohen, J., Nurcombe, V., Jeffrey, P. & Edgar, D. *Development* **107**, 381-387 (1989).
14. Keizer, G. D., Visser, W., Vliem, M. & Figdor, C. G. *J. Immun.* **140**, 1393-1400 (1988).
15. Bednarczyk, J. L. & McIntyre, B. W. *J. Immun.* **144**, 777-784 (1990).
16. Gulino, D., Ryckwaert, J.-J., Andrieux, A., Rabiet, M.-J. & Marguerie, G. *J. Biol. Chem.* **265**, 9575-9581 (1990).
17. Dustin, M. L. & Springer, T. A. *Nature* **341**, 619-624 (1989).
18. Plow, E. F. & Ginsberg, M. H. *Prog. Hemost. Thromb.* **9**, 117-156 (1989).
19. Figdor, C. G., van Kooyk, Y. & Keizer, G. D. *Immun. Today* **11**, 277-280 (1990).
20. Shimizu, Y., Van Seventer, G. A., Horgan, K. J. & Shaw, S. *Nature* **345**, 250-253 (1990).
21. Shaw, L. M., Messier, J. M. & Mercurio, A. M. *J. Cell Biol.* **110**, 2167-2174 (1990).
22. Yatogo, T., Izumi, M., Kashiwagi, H. & Hayashi, M. *Cell Struct. Funct.* **13**, 281-292 (1988).
23. Timpi, R. *et al. J. Biol. Chem.* **254**, 9933-9937 (1979).
24. Bodary, S. C., Napier, M. A. & McLean, J. W. *J. Biol. Chem.* **264**, 18859-18862 (1989).
25. Neugebauer, K. M., Emmett, C. J., Venstrom, K. & Reichardt, L. F. *Neuron* (in the press).

ACKNOWLEDGEMENTS. We thank K. Jones, M. Ignatius, C. Murphy-Erdosh and K. Tomaselli for their insights during this work; C. Damsky, C. Emmett, Z. Hall and F. Lefcort for comments on the manuscript; and C. Emmett and K. Venstrom for setting up the monoclonal screen. L.F.R. is an investigator of the Howard Hughes Medical Institute.

A deficiency of the homeotic complex of the beetle *Tribolium*

Jeffrey J. Stuart*, Susan J. Brown†, Richard W. Beeman & Robin E. Denell†‡

USDA, ARS, US Grain Marketing Research Laboratory, 1515 College Avenue, Manhattan, Kansas 66502, USA

† Division of Biology, Ackert Hall, Kansas State University, Manhattan, Kansas 66506, USA

IN *Drosophila*, the establishment of regional commitments along most of the anterior/posterior axis of the developing embryo depends on two clusters of homeotic genes: the Antennapedia complex (ANT-C) and the bithorax complex (BX-C). The red flour beetle has a single complex (HOM-C) representing the homologues of the ANT-C and BX-C in juxtaposition¹. Beetles *trans*-heterozygous for two particular HOM-C mutations spontaneously generate a large deficiency, presumably by an exchange within the common region of two overlapping inversions. Genetic and molecular results indicate that this deficiency spans at least the interval between the *Deformed* and *abdominal-A* homologues. In deficiency homozygous embryos, all gnathal, thoracic and abdominal segments develop antennal appendages, suggesting that a gene(s) has been deleted that acts to distinguish trunk from head. There is no evidence that beetles have a homologue of the segmentation gene *fushi tarazu* of similar genomic location and function. On the basis of the genetic tractability², convenient genome size and organization of *Tribolium*³, and its relatively long phylogenetic divergence from *Drosophila* (>300 million years), we have integrated developmental genetic and molecular analyses of the HOM-C. We isolated about 70 mutations in the complex representing at least six complementation groups. The homeotic phenotypes of adults² and lethal embryos lead us to believe that these beetle genes are homologous with the *Drosophila* genes indicated in Fig. 1.

Antennagalea (*Ag*) and *Abdominal-Extra sclerite 1* (*A^{Es1}*) are dominant HOM-C mutations. *A^{Es1}* suppresses all crossing over in the complex, whereas *Ag* is associated with a reduction in the rate of recovery of recombinants (data not shown). We observed that *Ag/A^{Es1}* heterozygotes spontaneously generate large, apparently identical deficiencies within the HOM-C (Table 1). Neither does so in other heterozygous combinations, and we presume that these deficiencies arise by an exchange within the common region of chromosomal inversions associated with each lesion⁴. Genetic characterization of each of five independently derived *Df*(HOM-C) chromosomes indicates that they are deleted at least for the interval from *Cephalothorax* (*Cx*) to *Abdominal* (*A*). First, in the heterozygous condition each expresses the haplo-insufficient phenotypes typical of the *Cx* and *A* loci. Second, each fails to complement mutant alleles at all known loci within this interval (Fig. 1). As the deficiencies complement mutations at *maxillopedia* (*mvp*) and *extra urogomphi* (*eu*), both endpoints must lie within the HOM-C.

Molecular experiments demonstrated that these are physical deletions and further defined their extent. We cloned the *Tribolium* homologue of *abdominal-A*, and showed that it corresponds to the genetically defined *Abdominal* gene (J.J.S. *et al.*, manuscript in preparation). An *Abdominal* probe recognizes restriction fragment length polymorphisms that were used to demonstrate that one of the spontaneous *Df*(HOM-C) chromosomes is deficient for that region (Fig. 2a). Although we have not yet identified any mutation in the beetle *Deformed* homologue, molecular cytogenetic studies show that it maps (as expected) to the left of *Cx* and is deleted by *Df*(HOM-C) (Figs 1 and 2b).

Embryos homozygous for *Df*(HOM-C) display a striking terminal lethal phenotype: the epidermal portions of the three gnathal segments (mandibular, maxillary and labial), the three thoracic segments, and abdominal segments 1–8 are transformed to resemble antennae (Fig. 3). Lethal embryos are greatly shortened with respect to wild type (presumably because the normal antennal segment elaborates relatively little cuticle), and occupy a position at the anterior end of the egg. Such anterior structures as the labrum, antennae, stomodaeum and eyespots develop normally, except that the latter often seem to be floating freely within the embryo. The urogomphi (sclerotized conical-shaped, terminal derivatives of A9) and the proctodaeum also appear relatively normal in those embryos that complete epidermal development; the urogomphi are not well formed on embryos that die earlier. The terminal flagellar segments of the homeotic antennae are often partially formed or missing. The homeotic antennae are otherwise usually well developed anteriorly but sometimes less so in the posterior abdomen, suggesting an anterior–posterior sequence in appendage differentiation which may be interrupted by embryonic death. A8 appendages that are well developed show a mixed segmental identity; they are clearly antennal proximally, but have urogomphus-like distal features.

Much of the multiantennae phenotype observed for *Tribolium Df*(HOM-C) homozygotes can be rationalized on the basis of the lethal phenotypes of the genes known to be deleted. Embryos homozygous for *A* mutations show a transformation of A2–A8 to resemble A1, although *Utx*¹ homozygotes develop a rudimentary thoracic leg on A1. Lethal *ptl* homozygotes show a strong transformation of the thoracic segments to antennae, and *Cx* alleles cause a transformation of labium to antenna as well as a cephalization of the first thoracic segment. Thus in embryos lacking all four functions, we would expect that the labium and thorax would be transformed to antenna owing to the lack of *Cx* and *ptl* functions, and that all abdominal segments would reiterate the antennal-like thorax because of a lack of *Utx* and *A* functions. But the deletion of these genes does not explain the transformation of the other two gnathal segments. The *Drosophila Deformed* gene is expressed in the maxillary and mandibular segments during normal development^{5,6}, and fly

TABLE 1 Isolation of deficiencies of the HOM-C

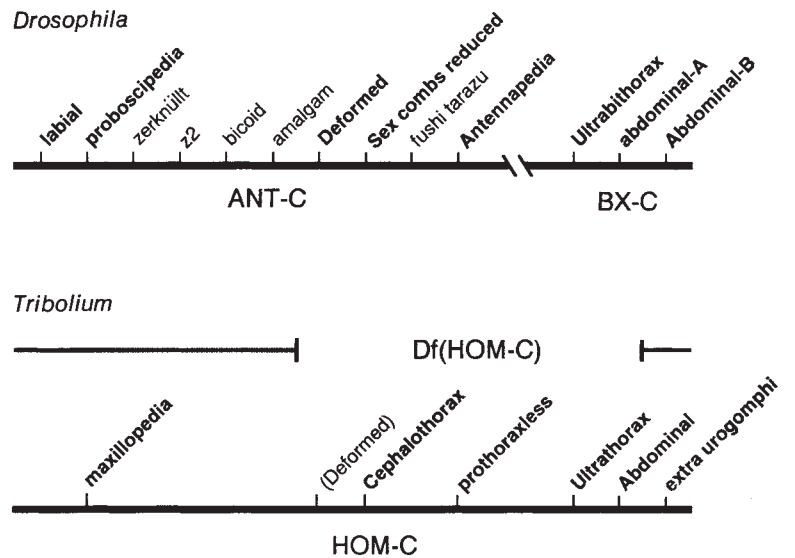
Treatment	Male genotype	Female genotype	Total progeny	No. putative deficiencies	No. studied
γ-rays	<i>Ag/A^{Es1}</i>	<i>s</i>	6,440	6	3
None	<i>Ag/A^{Es1}</i>	<i>A^{mas}</i>	8,333	11	2
None	<i>Ag/+</i>	<i>A^{mas}</i>	8,922	0	0
None	<i>A^{Es1}/+</i>	<i>A^{mas}</i>	6,775	0	0

Deficiency-bearing offspring were first observed in an experiment designed to isolate radiation-induced revertants of *Ag* and *A^{Es1}*. *A^{Es1}* is a dominant gain-of-function *Abdominal* mutation (J.J.S. *et al.*, manuscript in preparation) causing A2 to resemble A3. *Ag* transforms the base of the antenna to resemble that of the maxillary palp. The genetic significance of *Ag* is as yet unclear; it may be a gain-of-function variant at *maxillopedia* or represent a mutation at the *Tribolium Deformed* homologue. In the first experiment, males were treated with 4,000 roentgens of γ radiation as described², and the sooty (*s*) mutation (which is unlinked to the HOM-C) was used to facilitate separation of the sexes in order to discard irradiated males. Progeny heterozygous for putative deficiencies were identified on the basis of the lack of an *Ag* or *A^{Es1}* phenotype and expression of dominant phenotypes typically associated with *Cephalothorax* and *Abdominal* variants. In subsequent experiments, deficiencies were detected on the basis of their failure to complement the recessive mutant allele *A^{mas}*. The second experiment shows that deficiencies are generated spontaneously at a high frequency, and the third and fourth crosses demonstrate that they arise only when *Ag* and *A^{Es1}* are heterozygous with one another. Balanced stocks were established from all six individuals isolated in the first experiment and 2 of 11 isolated in the second; five of these lines were characterized.

* Present address: Department of Entomology, Purdue University, West Lafayette, Indiana 47907, USA.

‡ To whom correspondence should be addressed.

FIG. 1 The extent of *Df(HOM-C)* with respect to aligned maps of genes presently recognized in the homeotic complexes of *Drosophila*^{23,24} and *Tribolium*². Homeotic genes are printed in bold, and presumed homologues are printed on the same vertical. *Cephalothorax* and *prothoraxless* have not yet been ordered with respect to one another by recombination, nor have *Ultrathorax* and *Abdominal*; they are aligned here on the basis of their putative *Drosophila* homologues. Each *Df(HOM-C)* fails to complement the lethal alleles *Cx*⁶, *prothoraxless*^{D2} (*ptl*^{D2}), *Ultrathorax*¹ (*Utx*¹), and *A*¹⁰ with respect to viability. Each also fails to complement the visible phenotypes of the recessive viable alleles *Abdominal-missing abdominal sternites* (*A*^{mas}), *Abdominal-pointed abdominal sternites* (*A*^{pss}), and *Cephalothorax-alate prothorax* (*Cx*^{apt}). We confirmed that embryos bearing each lethal combination display the expected homeotic lethal syndrome. But the deficiencies all complement the recessive visible and lethal effects of *maxillopedia* (*mxp*) and *mxp*^{NO}, respectively, and the recessive visible phenotype of *extra urogomphi* (*eu*). Elsewhere, we have identified a duplication (*Dp(2)Dachs*) that arose as an aneuploid segregant from a chromosomal rearrangement associated with the *HOM-C* mutation *maxillopedia*^{Dachs}. This duplication complements the haplo-insufficient effects of *Cx*, *ptl* and *A* variants. Data shown in Fig. 2a demonstrate that *Dp(2)Dachs* is a physical duplication of the *Abdominal* gene. Figure 2b shows that the *Tribolium* *Deformed* homologue is deleted by *Df(HOM-C)*, and quantitative Southern hybridization confirms that result and shows that it is not included in the duplicated region (data not shown). The conclusions that *Cx* but not *Deformed* is duplicated (from genetic and molecular evidence, respectively) indicate that the beetle *Deformed* gene lies to the left of *Cx* as expected, and that the deficiency chromosomes are deleted at least for the region including *Dfd* through to *A*. Dashed lines indicate uncertainties as to the extent of the deficiency



with respect to possible homologues of *Drosophila* genes in the *zerknüllt*–*amalgam* interval.

embryos bearing any of a number of apparent null mutations are missing cuticular elements from these regions^{7,8}. Anterior transformations of the maxillae and mandibles of *Tribolium* embryos lacking the *Deformed* homologue may indicate that it does function as an embryonic homeotic selector gene in insects with more primitive cephalic organization.

We have evidence that *prothoraxless* is the homologue of the *Drosophila* *Antennapedia* gene (R.W.B., R. Garber and R.E.D., manuscript in preparation). The thorax → antennal transformation of *ptl*[−] beetle embryos is more similar to the phenotype of *Antp*[−] adult clones⁹ than *Antp*[−] embryos¹⁰, consistent with the idea that this phenotype reflects the primitive function of the gene⁹. These observations fit at least two evolutionary scenarios. Struhl⁹ suggested that thorax-like segments were converted to head segments by the evolution of new homeotic gene functions, and that *Antennapedia* arose to ensure that head-specific genes were repressed in the thorax. This view assumes the 'ground state' (the state of development in the absence of any homeotic selector gene function¹¹) to be trunk, although it should be noted that no homeotic selector gene has been recognized whose normal developmental function is to promote antennal rather than more posterior development. An alternative is that, as an ancient event preceding the origin of homeotic complexes, an *Antennapedia*-like function arose that acted positively in the trunk segments to distinguish them from the head. This picture, which postulates the ground state to be head, could have been concomitant with maternal mechanisms for establishing an anterior pole in the embryo, creating in turn the necessity to repress anterior development in the trunk. By duplication and divergence of function, this primordial *Antennapedia*-like gene could then have generated others in the homeotic complex that specify region-specific development within the trunk^{12,13}.

In *Drosophila*, the *fushi tarazu* (*ftz*) transcription unit lies between *Antennapedia* and *Sex combs reduced*¹⁴, and encodes a homeodomain protein that functions as a transcription factor¹⁵. Lethal embryos show a pair-rule phenotype in which even-numbered parasegments are deleted^{16,17}. Normal *ftz* expression depends on complex interactions with other segmentation genes¹⁸ and influences the expression of homeotic selector genes^{19,20}. In addition to its roles in the epidermis, *ftz* functions later in development in the central nervous system²¹.

If *Tribolium* were to have a *ftz* homologue of similar function and genomic location, it would be deleted by *Df(HOM-C)* (Fig.

1). The presence of appendages on fifteen gnathal and body segments of terminal homozygous embryos shows clearly that they express no pair-rule phenotype. This observation raises the question of whether *Tribolium* has a *ftz* homologue at all. Although we have cloned portions of six *Tribolium* genes that contain *Antennapedia*-class homeoboxes, we have not isolated a *ftz* homologue. Furthermore, no such homologue has been reported in any animal outside the Drosophilids²². Thus, it may be that *fushi tarazu* arose in the insect lineage subsequent to the divergence of beetles, bees and flies, indicating that its

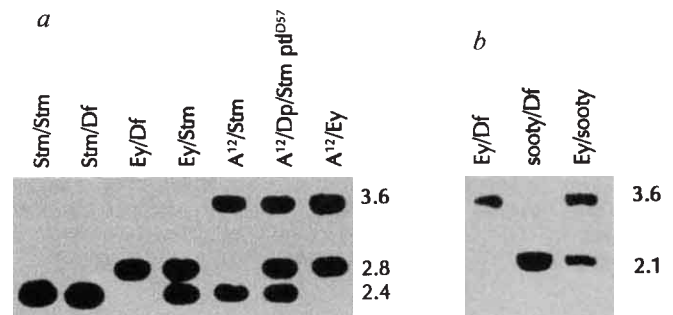


FIG. 2 Restriction fragment length polymorphism (RFLP) analysis of the extent of *Df(HOM-C)*. Genomic DNAs from the indicated genotypes were probed with *a*, a 2.5-kilobase (kb) cloned genomic fragment originating immediately 5' to the *Abdominal* homeobox (J.J.S. *et al.*, manuscript in preparation) or *b*, a 2.0-kb genomic fragment including the homeobox of the *Tribolium* *Deformed* homologue (S.J.B., unpublished observations). In *a*, the *Abdominal* probe detects an RFLP between *maxillopedia*–*Stumpy* (*mxp*^{Stm}) and *Eyeless* (*Ey*), two dominant mutations associated with chromosomal rearrangements that balance the entire *HOM-C*. The *Df(HOM-C)* chromosome tested has no genomic fragment complementary to this probe, as indicated by an examination of *mxp*^{Stm}/*Df(HOM-C)* and *Ey*/*Df(HOM-C)* heterozygotes. The *Abdominal* mutation *A*¹² is associated with a fragment of a unique size; the three bands detected in DNA from individuals heterozygous for *A*¹² and *mxp*^{Stm} and bearing *Dp(2)Dachs* demonstrate that an extra dose of the *Abdominal* gene is present. In *b*, the *Deformed* probe detects an RFLP between the stocks *Ey* and *sooty*; the mutation *sooty* is itself unlinked to the *HOM-C*. Again, the *Df(HOM-C)* chromosome is lacking a fragment complementary to this probe. In each case, *HindIII*-digested genomic DNA was resolved by agarose gel electrophoresis, blotted on to a nylon membrane and hybridized with gel-purified inserts ³²P-labelled by random priming. The sizes of hybridizing fragments are indicated to the right in kilobases.

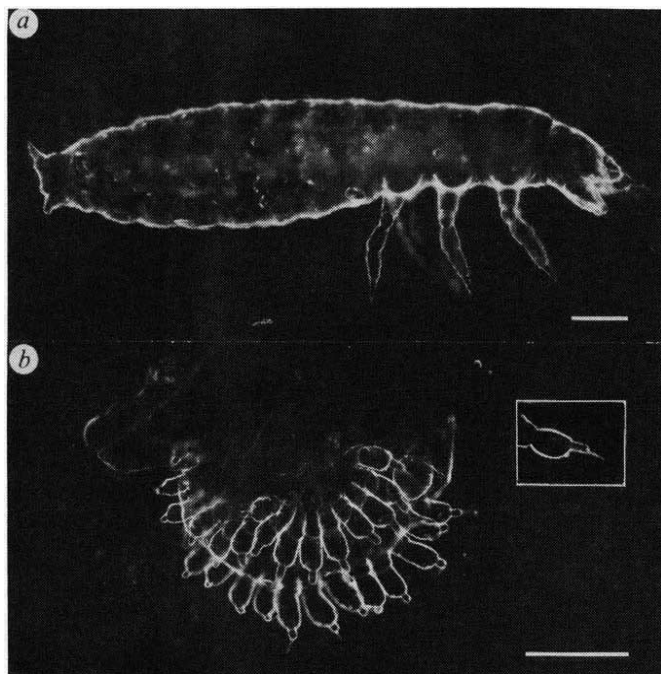


FIG. 3 A comparison of *a*, a newly hatched wild-type first instar larva and *b*, a lethal embryo (unhatched larva) homozygous for *Df(HOM-C)*. In the mutant larva, all gnathal, thoracic and abdominal segments develop antennal appendages or, for A8, an appendage with characteristics of both an antenna and urogomphus; insert in *b* shows a normal larval antenna for comparison. The mutant larva was dissected from the egg membranes, and lactic acid mounts¹¹ were prepared for each specimen. Each was photographed under dark-field illumination and is oriented with dorsal up and anterior to the right; the mutant larva is shown at a higher magnification; scale bars, 0.1 mm. The phenotypes of offspring bearing each of the independently derived *Df(HOM-C)* chromosomes were identical. Embryos with multi-antennal phenotypes arise only from crosses between *Df(HOM-C)/+* parents, and never from crosses of *Df(HOM-C)/+* beetles of either sex to wild-type mates. As they comprise less than the 25% expected from the former cross, it is likely that many deficiency homozygotes die during early embryonic development. Moreover, eggs generated by crosses between *Df(HOM-C)/+* parents die at a much higher rate than can be accounted for by the lethality of deficiency homozygotes, and egg hatch is also dramatically less than controls when heterozygotes of either sex are crossed to wild-type mates. Dissection of unhatched eggs from all crosses reveals larvae with a variety of abnormalities: mild dorsal fusions and missing denticles, curled or stubby tarsal claws, homeotic thoracic legs on A1, and extra urogomphi. It is likely that haplo-insufficient effects associated with the deficiency account for increased lethality and phenotypic abnormalities.

multiple regulatory interactions in *Drosophila* are of relatively recent evolutionary origin. □

Received 29 October 1990; accepted 7 January 1991.

1. Beeman, R. W. *Nature* **327**, 247–249 (1987).
2. Beeman, R. W., Stuart, J. J., Haas, M. S. & Denell, R. E. *Dev Biol.* **133**, 196–209 (1989).
3. Brown, S. J., Henry, J. K., Black, W. C. IV & Denell, R. E. *Insect Biochem.* **20**, 185–193 (1990).
4. Sturtevant, A. H. & Beadle, G. *Genetics* **21**, 554–604 (1936).
5. Chadwick, R. & McGinnis, W. *EMBO J.* **6**, 779–789 (1987).
6. Mahaffey, J. W., Diederich, R. J. & Kaufman, T. C. *Development* **105**, 167–174 (1989).
7. Merrill, V. K. L., Turner, F. R. & Kaufman, T. C. *Dev Biol.* **122**, 379–395 (1987).
8. Regulski, M., McGinnis, N., Chadwick, R. & McGinnis, W. *EMBO J.* **6**, 767–777 (1987).
9. Struhl, G. *Nature* **292**, 635–638 (1981).
10. Martinez-Arias, A. *EMBO J.* **5**, 135–141 (1986).
11. Lewis, E. *Nature* **276**, 141–152 (1978).
12. Akam, M., Dawson, I. & Tear, G. *Development* **104** (Suppl.), 123–133 (1988).
13. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. III. *Biochim. biophys. Acta* **989**, 25–48 (1989).
14. Weiner, A. J., Scott, M. P. & Kaufman, T. C. *Cell* **37**, 843–851 (1984).
15. Winslow, G. M., Hayashi, S., Krasnow, M., Hogness, D. S. & Scott, M. P. *Cell* **57**, 1017–1030 (1989).
16. Wakimoto, B. T. & Kaufman, T. C. *Dev Biol.* **81**, 51–64 (1981).
17. Nüsslein-Volhard, C., Kluding, H. & Jürgens, G. *Cold Spring Harb. Symp. quant. Biol.* **50**, 145–154 (1985).
18. Carroll, S. B. *Cell* **60**, 9–16 (1990).
19. Duncan, I. *Cell* **47**, 297–309 (1986).
20. Ingham, P. W. & Martinez-Arias, A. *Nature* **324**, 592–597 (1986).

21. Doe C. Q., Hiromi, Y., Gehring, W. J. & Goodman, C. S. *Science* **239**, 170–175 (1988).
22. Walldorf, U., Fleig, R. & Gehring, W. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9971–9975 (1989).
23. Mahaffey, J. W. & Kaufman, T. C. in *Developmental Genetics of Higher Organisms: A Primer in Developmental Biology* (ed. Malacinski, G. M.) 329–360 (Macmillan, New York, 1989).
24. Duncan, I. A. *Rev. Genet.* **21**, 285–319 (1987).

ACKNOWLEDGEMENTS. We thank K. Hummels and S. Haas for technical assistance, and T. Kaufman and W. Bender for comments. This work was supported by the American Cancer Society, National Aeronautics and Space Administration, National Science Foundation, US Department of Agriculture, and Wesley Foundation.

Pituitary hyperplasia and gigantism in mice caused by a cholera toxin transgene

Frank H. Burton*, Karl W. Hasel*, Floyd E. Bloom† & J. Gregor Sutcliffe*

Departments of * Molecular Biology and † Neuropharmacology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

CYCLIC AMP¹ is thought to act as an intracellular second messenger, mediating the physiological response of many cell types to extracellular signals^{2,3}. In the pituitary, growth hormone (GH)-producing cells (somatotrophs) proliferate and produce GH in response to hypothalamic GH-releasing factor^{4–6}, which binds a receptor that stimulates G_s protein activation of adenyl cyclase^{3,9–12}. We have now determined whether somatotroph proliferation and GH production are stimulated by cAMP alone^{5,7,11,13–15}, or require concurrent, non-G_s-mediated induction of other regulatory molecules by designing a transgene to induce chronic supraphysiological concentrations of cAMP in somatotrophs. The rat GH promoter^{16,17} was used to express an intracellular form of cholera toxin¹⁸, a non-cytotoxic and irreversible activator of G_s (ref. 19). Introduction of this transgene into mice caused gigantism, elevated serum GH levels, somatotroph proliferation and pituitary hyperplasia. These results support the direct triggering of these events by cAMP, and illustrate the utility of cholera toxin transgenes as a tool for physiological engineering.

Polymerase chain reaction²⁰ was used to amplify the 582-base-pair (bp) portion of the operon for *Vibrio cholerae* holotoxin¹⁸ that encodes its enzymatically active intracellular subunit, A1, while adding a eukaryotic translation initiator²¹ and terminator. This cholera toxin gene cassette was inserted into the 5' untranslated region of the rat GH-human GH fusion gene¹⁷, a construct used previously in transgenic mice to express human GH or diphtheria toxin (which ablates cells²²) specifically in pituitary somatotrophs¹⁷. The resulting 'rGH-CT' gene was used to make transgenic mice (Fig. 1).

Founders, F₁ transgenics, and non-transgenic littermates were weighed and endogenous serum GH (mGH) levels measured from a representative sample (Table 1). Of 13 founders, nine had markedly higher serum mGH levels than controls and eight had greater body weights. The average body weight of 46 F₁ transgenics was 1.3 times greater ($P < 0.001$) than that of 49 sex-matched littermate controls. The average serum mGH level of 35 F₁ transgenics (558 ng ml⁻¹) was over 21-fold higher ($P < 0.001$) than that of 21 F₁ controls (<26 ng ml⁻¹), with individual values as high as 2,940 ng ml⁻¹. Serum levels of two other cAMP-responsive pituitary hormones, prolactin and thyroid-stimulating hormone (TSH), were not elevated in transgenics relative to controls (data not shown), consistent with the transgene's action being somatotroph-specific. The mGH levels in transgenic mice remained elevated long into adulthood rather than abating after adolescence (Table 1), suggesting that both the mGH gene and the rGH-CT transgene are transcriptionally stimulated by cAMP in response first to GH-releasing factor and thereafter to cholera toxin, and that there may be no

physiological mechanism for repressing GH transcription other than by reducing cAMP levels.

Transgenic pituitaries were markedly larger than control pituitaries (Fig. 2a), with 'ballooning' of anterior lobes. Although the average control pituitary weight was 1.9 mg (s.d. = 0.4 mg), pituitaries from three transgenics with highly elevated serum mGH levels (848 to 2,940 ng ml⁻¹) weighed 7 mg, 9 mg and 9.3 mg, respectively. The average pituitary weight from eight transgenics with more moderately elevated serum GH levels (336 to 736 ng ml⁻¹) was 4.5 mg (s.d. = 0.8 mg).

Hyperplasia of the transgenic pituitaries could be due to cholera toxin-induced proliferation of either somatotrophs or both somatotrophs and mammosomatotrophs, which are somatotrophs developing into prolactin-releasing cells and which transiently release both hormones²³. We used immunocytochemical stains for cholera toxin, GH, prolactin and TSH (a pituitary hormone whose producer cells would not be expected to express or be influenced by the transgene) to examine transgenic and control anterior pituitaries (posterior and intermediate pituitaries were negative with all four immunostains; data not shown). In areas of transgenic pituitaries where GH-positive cells were not uniformly distributed, cholera toxin reactivity co-localized with the GH-positive regions, but not with prolactin- or TSH-positive regions (Fig. 2d-g). In

control male pituitaries, 53% (s.d. = 5.8%) of the cells were GH-reactive, compared with 79% (s.d. = 2.9%) in transgenics (Fig. 2b and c). Similar results were seen in transgenic females (not shown). Prolactin- or TSH-positive cells in male transgenics showed no change in absolute number, and thus a marked decrease in density (Fig. 2h-j and data not shown). The increase in somatotroph to non-somatotroph ratio from ~1:1 to 4:1, with no change in the numbers of other cell types, requires a 2.5-fold increase in anterior pituitary size, consistent with the observed 2.4-fold increase in average transgenic pituitary weight. These data suggest that cholera toxin expression is restricted to somatotrophs and that selective proliferation of only these cells causes the hyperplasia.

The markedly lower density of prolactin-positive cells in male transgenics than in controls (Fig. 2h,j) indicates that elevated cAMP is not converting the amplified somatotroph population into mammosomatotrophs. If GH-releasing factor triggers this conversion²⁴, it must be acting through another mechanism. Transgenic females have many more prolactin-positive cells than males (Fig. 2i,j), perhaps due to female-hormone stimulation of the conversion by a non-cAMP-mediated pathway.

Pituitaries of 24 transgenics and 43 controls aged 7–13 months were examined for tumours by gross anatomy, with 13 from each group examined further by histology. None exhibited a

TABLE 1 GH levels and growth of rGH-CT transgenic mice

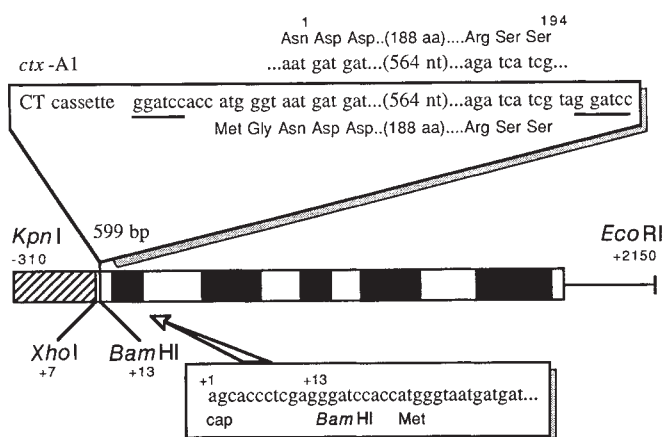
Mouse	Transgenic (+)	Serum mGH (ng ml ⁻¹)	Growth (ratio)	Mouse	Transgenic (+)	Serum mGH (ng ml ⁻¹)	Growth (ratio)
F₀1 ♀	+	848	1.43	F₀52 ♂	+	339	0.96
F ₁ 3 ♂		44		F ₁ 55 ♂	+	108	1.27
F ₁ 4 ♂	+	580	1.33	F ₁ 56 ♂		16	
F ₁ 5 ♂	+	1,430	1.22	F ₁ 57 ♂	+	336	1.08
F ₁ 6 ♂		67		F ₁ 58 ♀	+	87	1.12
F ₁ 2 ♀	+	617	ND	F ₁ 59 ♀	+	78	1.28
F ₁ 7 ♀	+	1,090	ND	F ₁ 60 ♀		<7	
F₀8 ♀	+	484	1.22	F ₁ 61 ♀	+	505	1.42
F ₁ 8 ♂	+	545	1.53	F ₁ 62 ♀	+	250	1.58
F ₁ 9 ♂	+	408	1.31	F₀62 ♂	+	753	1.07
F ₁ 10 ♂	+	226	1.17	F ₁ 64 ♂	+	736	1.07
F ₁ 11 ♂		15		F ₁ 66 ♂		16	
F ₁ 12 ♂	+	868	1.53	F ₁ 73 ♂		<7	
F ₁ 14 ♀		8		F ₁ 74 ♂		<7	
F ₁ 15 ♀	+	505	1.55	F ₁ 75 ♂	+	292	1.17
F₀15 ♂	+	639	1.29	F ₁ 76 ♂	+	155	1.39
F ₁ 16 ♂	+	413	1.47	F ₁ 67 ♀		42	
F ₁ 24 ♂		121		F ₁ 68 ♀	+	279	1.12
F ₁ 17 ♀	+	143	1.16	F ₁ 69 ♀	+	387	1.27
F ₁ 19 ♀		12		F ₁ 70 ♀	+	456	1.34
F ₁ 21 ♀	+	361	1.55	F ₁ 72 ♀		76	
F ₁ 22 ♀	+	421	1.24	F₀65 ♀	+	466	1.22
F ₁ 23 ♀		17		F ₁ 79 ♂	+	1,980	ND
F₀19 ♂	+	432	1.34	F ₁ 80 ♂	+	2,940	ND
F ₁ 42 ♀		43		F ₁ 81 ♂	+	292	ND
F ₁ 43 ♀	+	691	1.16	F ₁ 83 ♀	+	80	1.08
F ₁ 45 ♀		13		F ₁ 84 ♀		<7	
F₀44 ♀	+	371	1.22	F ₁ 85 ♀	+	369	1.27
F ₁ 46 ♂		<7		F ₁ 86 ♀		9	
F ₁ 47 ♂		<7		F ₁ 87 ♀	+	26	1.18
F ₁ 48 ♂	+	932	1.47	F₀33 ♂	+	44	0.97
F ₁ 49 ♀		<7		F₀35 ♀	+	10	0.91
F ₁ 52 ♀	+	538	1.34	F₀36 ♀	+	462	1.12
F ₁ 54 ♀	+	415	1.19	F₀37 ♀	+	7	0.98
				F₀43 ♀	+	56	1.11

Founder transgenic mice (F₀) are shown in bold type. F₁ progeny of individual founders are listed in indented columns below each founder. Serum mGH levels were measured by anti-rat GH radioimmunoassay performed on duplicate samples by Hazleton Biotechnologies. The minimum threshold of detectability by this assay was 7 ng ml⁻¹. Growth ratios were determined for F₀ mice at 2 months of age and for F₁ mice at 5 to 7 months of age. Serum mGH levels were determined for F₀ mice at 10 months of age and for F₁ mice at 5.5 to 7.5 months of age. There was no correlation between transgene copy number (data not shown) and body weight or mGH levels, except for mice suspected of having acquired only partial integrations: of the four founders with a transgene copy number of 1 or less (F₀15, F₀33, F₀37, F₀43), three did not express elevated mGH levels. No human GH was detected in the sera of transgenics (data not shown), supporting the idea that cholera toxin cassette insertion into the 5' untranslated region of rGH-hGH prevents translation of the downstream human GH open reading frame.

ND, Not determined; growth ratios were not calculated because there were no control littermates of the same sex.

FIG. 1 Construction of a rat growth hormone/cholera toxin (rGH-CT) fusion transgene. Top: amino-acid and nucleotide sequence of the portion of the *V. cholerae* *ctx* operon encoding the cholera toxin A1 (ref. 18) subunit, aligned with that of a polymerase chain reaction-amplified mouse CT cassette. Inclusion of the consensus eukaryotic translation initiator²¹ 'CCAC-CATGG' adds two N-terminal amino acids to the sequence of mouse cholera toxin. Polymerase chain reaction primers used to create the cassette were CT-5 (5'-GCGGATCCACCATGGGTAATGATGATAAGTTATAT-3') and CT-3 (5'-GCGGATCCTACGATGATCTTGAGCATTC-3'). *Bam*HI cloning sites are underlined. Middle: ligation of the 599-bp CT cassette into the *Bam*HI site at position +13 within the 5' untranslated region of the rat growth hormone/human growth hormone (rGH-hGH) fusion gene¹⁷. Components of the rGH-hGH fusion gene: shaded box (rGH promoter and 5' untranslated region); open boxes (hGH 5' untranslated region, introns, and 3' untranslated region); solid boxes (hGH coding regions). Insertion of the cassette converts all human sequences into 3' untranslated sequence. Bottom: sequence of the 5' untranslated region of rGH-CT.

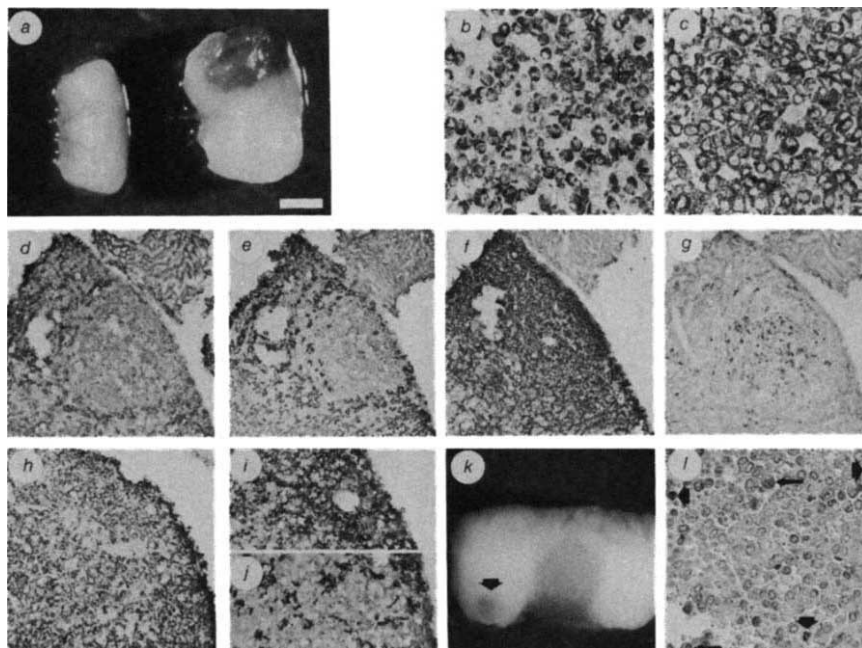
METHODS. The rGH-CT structural gene, carried on a 3-kb *Kpn*1-*Eco*RI fragment, was excised from pUC vector DNA for microinjection at 700 molecules per pronucleus into 200 BALB/c × C57BL/6 hybrid eggs^{26,27}. Sixty-eight animals developed from the microinjected eggs. Hybridization of radiolabelled CT cassette DNA to Southern-blotted tail DNA prepared at weaning revealed that 13 of the mice carried the transgene, nine with two or more



copies. Breeding of eight transgenic lines yielded 97 F_1 generation mice, 47 of which were transgenic. Southern blotting of DNA from these eight lines identified 10 transgenic loci, of which eight showed 50% inheritance.

FIG. 2 Pituitary and somatotroph hyperplasia in rGH-CT transgenic mice. *a*, Whole pituitaries of control (left) and transgenic (right) mice. Pituitary weights are 2.1 mg and 9.2 mg, respectively. The lesion on the transgenic pituitary is an infarct, perhaps caused by organ compression occurring from hyperplasia. *b-j*, Immunohistochemical localization of cholera toxin (CT)-, GH-, prolactin (PRL)- or TSH-positive cells within control (*b, h*) or transgenic (*c, d-g, i-j*) mouse pituitaries: *b* and *c*, GH staining shows greater density of GH-positive cells in transgenics (*c*); *d-g*, Contiguous sections show colocalization of CT (*d*) and GH (*e*) staining but not PRL (*f*) or TSH (*g*) staining; *h-j*, PRL staining shows, relative to controls of either sex (*h*), a decreased density of PRL-positive cells in male (*j*) but not female (*i*) transgenics. *k*, A noninfarcted raised symmetrical lesion (arrow) in the anterior lobe of a transgenic male; *l*, Orange G-PAS staining of the same lesion showing mitotic figures (large arrows), binucleate cells (small arrows) and cavities containing blood cells. The pituitary cells in the lesion are uniformly GH-positive and acidophilic (not shown), and exhibit a 20-fold increased mitotic index (3 mitotic figures/field versus 0.15 mitotic figures/field outside the lesion), but are not encapsulated.

METHODS. Pituitaries in *d-j* were isolated after perfusion with 4% paraformaldehyde and cryostat sectioned. Pituitaries in *b, c, k* and *l* were isolated without perfusion, fixed in 10% buffered formalin, and paraffin sectioned. All sections were immunoperoxidase stained with these antisera concentrations: anti-rGH (1:2,500), anti-CT (1:1,000), anti-mPRL (1:500), anti-rTSH β (1:500). Cell counts using light haematoxylin counterstain (not shown) were verified on



histochemically stained sections. Number of control and transgenic pituitaries counted: GH, $n=4$; Orange G-periodic acid-Schiff, $n=3$; PRL, $n=3$; TSH, $n=8$. Scale bar for all panels is shown in *a*, and is 1 mm for *a*, 0.55 mm for *k*, 190 μ m for *d-g*, 120 μ m for *h-j* and 30 μ m for *b, c* and *l*.

clearly defined tumour, although a transgenic pituitary exhibited a GH-positive hyperplastic focus (Fig. 2*l*) seen anatomically as a raised, non-infarcted circular lesion on an anterior lobe that otherwise exhibited generalized hyperplasia (Fig. 2*k*). The lack of clearly defined tumours suggests either that elevated cAMP concentrations do not facilitate their development or, given the later onset of such tumours in transgenic mice with elevated serum GH-releasing factor levels⁶, that a longer period of somatotroph proliferation is required. Secondary oncogenic events other than elevation of cAMP might also be needed.

The cholera toxin transgene has shown that elevation of intracellular cAMP concentrations is the probable cause of the somatotroph's response to GH-releasing factor—stimulation of mGH production and cell proliferation. The transgene was efficiently expressed, producing the predicted phenotype from 11 of 15 transgenic loci (Fig. 1). As with other ADP-ribosylating

toxins, just 4 to 10 cholera toxin molecules per cell can maximally alter cell activity²⁵. Thus one might expect that cholera toxin expression from non-cAMP-inducible or weak promoters would still be sufficient to alter cAMP concentrations. Cholera toxin transgenes should be useful tools for studying the consequences of aberrantly elevated cAMP concentrations *in vivo* in any cell type that can be defined by a specific promoter, and could lead to the creation of mouse models for many human diseases. □

Received 16 August; accepted 17 December 1990.

- Sutherland, E. W. & Rall, T. W. *J. Am. chem. Soc.* **79**, 3608 (1957).
- Robison, G. A., Butcher, R. W. & Sutherland, E. W. *A. Rev. Biochem.* **37**, 149-174 (1968).
- Neer, E. J. & Clapham, D. E. *Nature* **333**, 129-134 (1988).
- Asa, S. L. *et al. J. Clin. Endocr. Metab.* **58**, 796-803 (1984).
- Barinaga, M., Bilezikjian, L. M., Vale, W. W., Rosenfeld, M. G. & Evans, R. M. *Nature* **314**, 279-281 (1985).

6. Hammer, R. E., Brinster, R. L., Rosenfeld, M. G., Evans, R. M. & Mayo, K. E. *et al. Nature* **315**, 413–416 (1985).
7. Billestrup, N., Swanson, L. W. & Vale, W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6854–6857 (1986).
8. Asa, S. L. *et al. Proc. Soc. exp. Biol. Med.* **193**, 232–235 (1990).
9. Bilezikjian, L. M. & Vale, W. *Endocrinology* **113**, 1726–1731 (1983).
10. Smigel, M. *et al. Adv. cyclic Nucl. Prot. Phosphoryl. Res.* **17**, 1–18 (1984).
11. Vallar, L., Spada, A. & Giannattasio, G. *Nature* **330**, 566–568 (1987).
12. Struthers, R. S., Perrin, M. H. & Vale, W. *Endocrinology* **124**, 24–29 (1989).
13. Schofield, J. G. *Nature* **215**, 1382–1383 (1967).
14. Bourne, H. R. *Nature* **330**, 517–518 (1987).
15. Landis, C. A. *et al. Nature* **340**, 692–696 (1989).
16. Barta, A., Richards, R. I., Baxter, J. D. & Shine, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4867–4871 (1981).
17. Behringer, R. R., Mathews, L. S., Palmiter, R. D. & Brinster, R. L. *Genes Dev.* **2**, 453–461 (1988).
18. Mekalanos, J. J. *et al. Nature* **306**, 551–557 (1983).
19. Gill, D. M. & Meren, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3050–3054 (1978).
20. Saiki, R. K. *et al. Science* **239**, 487–491 (1988).
21. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
22. Palmiter, R. D. *et al. Cell* **50**, 435–443 (1987).
23. Hoeffler, J. P., Boockfor, F. R. & Frawley, L. S. *Endocrinology* **117**, 187–195 (1985).
24. Stefanescu, L. *et al. Endocrinology* **125**, 2710–2718 (1989).
25. Middlebrook, J. L. & Dorland, R. B. *Microbiol. Rev.* **48**, 199–221 (1984).
26. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. & Ruddle, F. H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7380–7384 (1980).
27. Brinster, R. L. *et al. Cell* **27**, 223–231 (1981).

ACKNOWLEDGEMENTS. We thank F. Dixon and N. Sarvetnick for reading the manuscript, J. Mekalanos for the *ctx* operon and anti-CT antisera, R. Palmiter for the rGH-hGH plasmid, A. F. Parlow, the NIDDK and the National Hormone & Pituitary Program, University of Maryland School of Medicine, for anti-rGH, anti-mPRL and anti- β TSH antisera, J. Price, E. Kothari, P. Charos, D. Alvarez, E. Battenberg, and M. Pagels for technical assistance, and L. Elder for preparing the manuscript. This work was supported by the NIH, the American Cancer Society (F.H.B.), the Fonds de la Recherche en Santé du Québec and the Alzheimer's Disease and Related Disorders Association (K.W.H.).

Restoration by a 35K membrane protein of peroxisome assembly in a peroxisome-deficient mammalian cell mutant

Toshiro Tsukamoto, Satoshi Miura & Yukio Fujiki*

Laboratory of Molecular Cell Biology, Meiji Institute of Health Science, Odawara, Kanagawa 250, Japan

PEROXISOMES are among the intracellular organelles of eukaryotic cells that contain specialized sets of enzymes with specific functions¹. Little is known of membranous components involved in assembly of the intracellular compartments^{2–5}. We isolated two peroxisome-deficient and mutually complementary, Chinese hamster ovary cell mutants, Z65 and Z24⁶, which closely resembled fibroblasts from patients with autosomal recessive, peroxisome-defective disorders such as Zellweger syndrome^{1,7}. These patients show characteristic dysmorphism, severe hypotonia, psychomotor retardation, and peroxisomal dysfunctions and rarely survive early childhood. Here we report what seems to be the first direct cloning and characterization of a complementary DNA encoding a peroxisomal membrane protein of relative molecular mass 35,000 (*M_r* 35K) that restores the biogenesis of peroxisomes and complements the defect of peroxisomal functions in the mutant Z65.

The mutants Z65 and Z24 seem to be defective in the assembly of peroxisomes, whereas the biosynthesis of peroxisomal proteins seems normal⁶. We searched for a gene encoding a factor that complements the defects of the mutant Z65, by transfection of a rat liver cDNA library constructed in a mammalian expression vector, pcD2⁸. Revertant cells were selected by the P12/UV (12-(1'-pyrene)dodecanoic acid/ultraviolet) method⁹, where ~90% of the wild-type Chinese hamster ovary (CHO) cells survived but only 0.02% or less of the Z65 were viable, presumably due to a hypersensitivity to the reactive oxygen species formed by irradiation of P12 (Table 1c). P12/UV-resistant cells were obtained in two experiments (Table 1a). In experiment 1, peroxisomes were absent in the resistant cells. The decrease in uptake of P12 by these cells or the increase in radical-scavenging activity may provide P12/UV resistance

TABLE 1 Isolation and characterization of peroxisome-restored revertant cells

(a) Isolation of P12/UV-resistant revertant of CHO cell mutant Z65

Expt	Total G418 ^R transformants*	P12/UV selection† (%)		Peroxisome-positive cells‡ (%)
		1st	2nd	
1	2.9 × 10 ⁴	0.160	28	0
2	3.8 × 10 ⁴	0.0034	29	>95

(b) Complementing activity of recovered plasmids (*Bam*HI fragments of each plasmid)

Plasmid	Length of fragments (kb)	Peroxisome-positive clone
pcD65A	0.4, 0.5, 0.9	0/20
pcD65B	0.56, 0.6, 2.9§	5/20
pcD65C	0.13, 0.35, 1.3§	12/20
pcD65D	0.6§, 1.9§	15/20
pcD65E	1.2, 1.6	0/20
pL2	—	0/20

(c) Properties of wild-type, mutant Z65, and PAF-1-transfected CHO cells

Cell	Peroxisomes	Catalase latency (%)	DHAP-ATase (%)	P12/UV (%)
CHO	+	67	100	87
Z65	—	0	<1	0.02
TR1	+	46	110	92
TD1	+	70	125	88

Methods: a, Poly(A)⁺RNA^{22,23} was prepared from a female Sprague-Dawley rat that had been treated with clofibrate, a hypolipidaemic drug²⁴. A cDNA library, containing 2.5 × 10⁵ independent clones, was constructed in the expression vector pcD2⁸, except that oligo(dT)-tailed pcDV1-PL (Pharmacia) was used as a vector-primer instead of pcDV1. 6-thioguanine-resistant variant cells (5 × 10⁵) of the mutant Z65⁶ were cultured in each of three 100-mm dishes and transfected with 15 µg DNA⁸. Cells were put into 30 dishes, grown for 5–6 days in the presence of 400 µg ml⁻¹ G418, harvested and inoculated into three (experiment 1) or five (experiment 2) 100-mm dishes at a density of 5 × 10⁵ cells per plate. P12/UV selection was performed at 2 µM P12 for 16 h followed by 3 min ultraviolet irradiation according to Zoeller *et al.*⁹ After 4–6 days, all surviving cells were reinoculated into two 100-mm dishes and subjected to a second P12/UV selection. The cells were collected and some were then assayed for the presence of peroxisomes (Fig. 2a). In experiment 2, the P12/UV resistant cells were inoculated at a lesser density, to isolate clones as a single colony. Six of seven isolated colonies were peroxisome-positive, one of which, TR1, was further analysed. b, Plasmids were recovered by means of polyethylene glycol-mediated cell fusion⁷ between P12/UV-resistant cells (not cloned) and COS-7 cells¹⁰. Plasmids were prepared from 12 independent *Escherichia coli* colonies, which contained five different clones (pcD65A to E). These five plasmids and a control plasmid pL2⁸ were used for transfection assay as described above, except that a 60-mm dish was used. Of the transfected cells, 1/10 or 1/20 was inoculated into 60-mm or 100-mm dishes, respectively. Cells in the 60-mm dish were kept in G418-containing medium for 2 days, 1/10 was further cultured in the same medium for 3 days in a 35-mm dish; peroxisome-positive colonies were identified by immunocytochemical staining. The cells in a 100-mm dish were grown for 10 days, before isolating the transfectants as a single colony. Three of four pcD65D-transfectants were peroxisome-positive, one of which, TD1, was analysed. c, Morphological analysis was performed as in Fig. 2a. Catalase latency, measured at 100 µg ml⁻¹ digitonin, was expressed as a percentage (100%, relative free catalase activity⁶). Dihydroxyacetonephosphate acyltransferase (DHAP-ATase) was measured as described⁶ and presented as a percentage relative to that of wild-type CHO cells (88 pmol min⁻¹ mg⁻¹). Protein was determined by dye-binding assay (Bio-Rad). For determination of P12/UV resistance, 200 or 1 × 10⁵ cells were inoculated into 60-mm dishes and selected as above.

* Percentage of G418^R colonies was ~0.3% in both experiments.

† The survival rate after P12/UV selection was expressed as a percentage of the unselected control.

‡ Percentage of peroxisome-positive cells was estimated, relative to the resistant cells after the second P12/UV selection.

§ Subcloned fragments into pUC18 as in Fig. 1a.

|| A plasmid containing neomycin-resistant gene but not cDNA, as a negative control.

* To whom correspondence should be addressed.

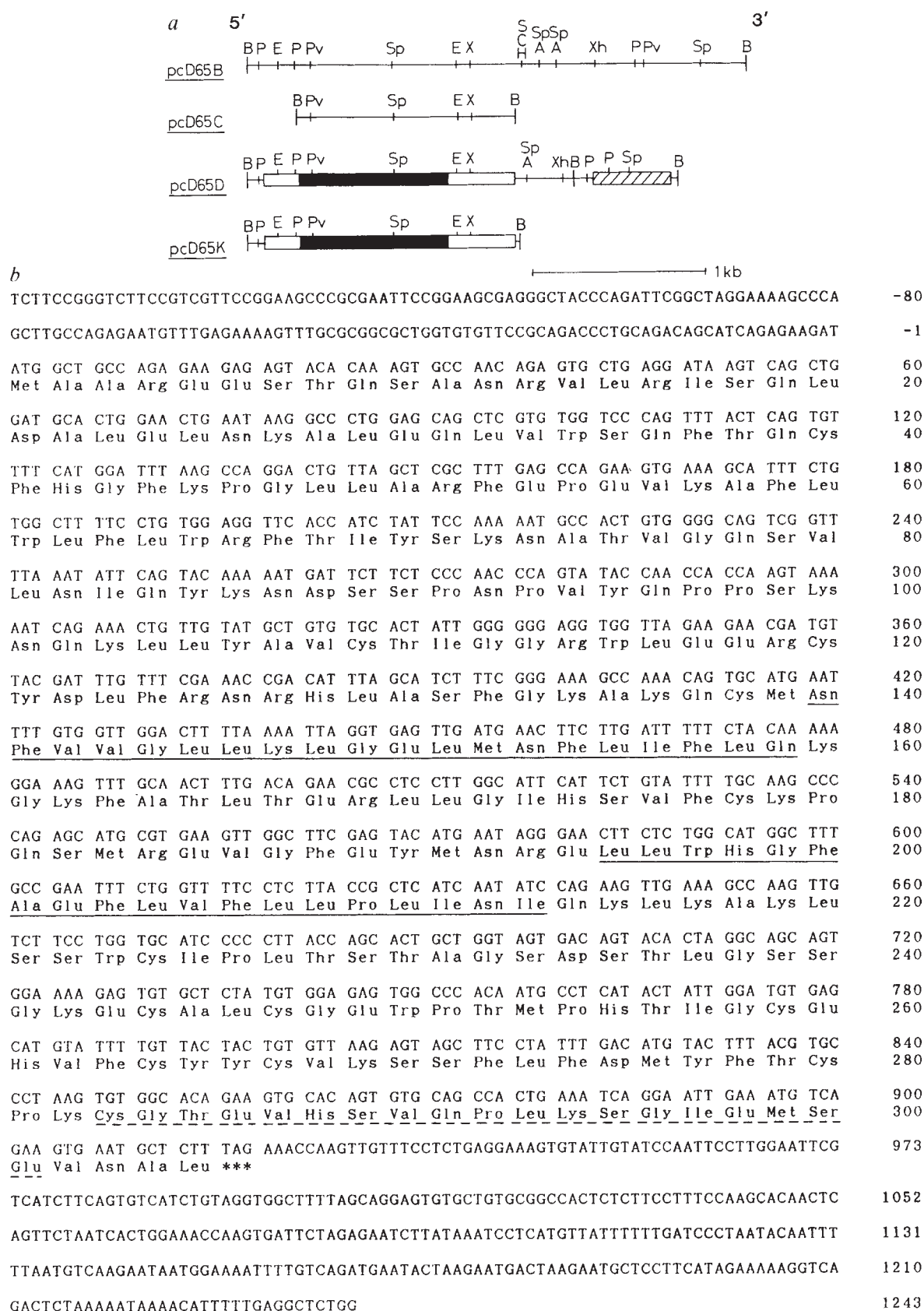


FIG. 1 Complementary DNA coding for peroxisome assembly factor-1 (PAF-1). **a**, Restriction enzyme map of subcloned cDNA fragments. A, *Avall*; B, *Bam*HI; C, *Clat*; E, *Eco*RI; H, *Hind*III; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I; Sp, *Sph*I; X, *Xba*I; Xh, *Xho*I. Open and filled boxes indicate non-coding and coding regions of PAF-1 cDNA, respectively; diagonal hatched box shows smaller cDNA. **b**, Nucleotide and deduced amino-acid sequence of PAF-1 cDNA. Two putative hydrophobic membrane-spanning regions are underlined; the dashed line, the sequence used for chemical synthesis of peptide.

METHODS. **a**, *Bam*HI fragments of cDNA were subcloned into pUC18, and digested with various restriction enzymes. **b**, Overlapping deletions in both orientations of PAF-1 cDNA clones in pUC18 were prepared with a Kilo-sequencing Deletion kit (Takara Shuzo, Kyoto). Nucleotide sequence was determined using Sequenase (USB) with 7-deaza-dGTP. Hydrophathy analysis¹⁶ was done using the algorithm with a window ranging 19 amino acids. Sequences have been submitted to EMBL.

without the formation of peroxisomes. In the second experiment almost all the resistant cells contained peroxisomes. Plasmids (pcD65A to E) were recovered from these cells¹⁰, three (B, C and D) of which exhibited peroxisome-restoring activity (Table 1b). The largest *Bam*HI fragment of each plasmid (Table 1b) contained a 1.0-kilobase (kb) common *Pvu*II-*Xba*I fragment (Fig. 1a), implying that these plasmids were likely to be generated from the same plasmid by recombination during recovery of the plasmid¹⁰. Restriction enzyme mapping of pcD65B suggested that the pcD2 vector was partially duplicated. The plasmid pcD65D was revealed, by sequencing two fragments (Table 1b), to contain two different cDNAs; a large portion of the open reading frame (Fig. 1a, filled box) of the longer cDNA was present in the *Pvu*II-*Eco*RI fragment which was commonly found in the three biologically active clones. The smaller cDNA (Fig. 1a, hatched box) proved to encode the β subunit of rat propionyl-CoA carboxylase and was absent in pcD65C, as determined by Southern blotting (data not shown). We conclude that the longer cDNA possesses complementing activity for peroxisome-deficient Z65 cells and encodes the 35K peroxisomal membrane protein (termed peroxisome assembly factor-1 or PAF-1). In addition, pcD65K obtained by rescreening of the original cDNA library contained only the same PAF-1 cDNA as in pcD65D (Fig. 1a) and likewise complemented the Z65. There is no homology of the PAF-1 sequence to any known protein, thereby implying that PAF-1 is a newly identified 35K protein of 305 amino acids (Fig. 1b).

We then examined whether PAF-1 would restore several biochemical abnormalities in the mutant Z65. Peroxisomes were detected by immunocytochemical light microscopy (Fig. 2a) and cytochemical electron microscopy (data not shown), in Z65·TR1 (TR1), a revertant isolated from transfectants of a cDNA library, and Z65·TD1 (TD1), a peroxisome-restored transfectant of pcD65D. In TR1 and TD1, catalase appeared to be localized in peroxisomes, as determined by its latency (Table 1c). Acyl-CoA oxidase, the first enzyme of the peroxisomal fatty-acid β -oxidation system, is a heterodimer comprising 75K(A), 53K(B) and 22K(C) components^{6,11}; B and C are derived from A^{5,12}. All components were present in TR1 and TD1 cells as in the wild type, whereas only the greatly reduced A form was seen in Z65, because of a rapid degradation⁶ (Fig. 2b, lanes 1-4). The third enzyme, 3-ketoacyl-CoA thiolase, was evident as a mature 41K form in TR1, TD1, and wild-type cells, whereas only a 44K precursor was noted in Z65⁶ (lanes 5-8). Peroxisomal dihydroxyacetonephosphate acyltransferase, catalysing the first step of plasmalogen synthesis¹, was restored in its activity in both TR1 and TD1 (Table 1c). Moreover, TD1, not selected by P12/UV, showed resistance, as noted for wild-type and TR1 cells. Accordingly, the 35K PAF-1 is most likely to be sufficient to complement defects of the mutant. But the clone pcD65D did not rescue another CHO mutant, Z24 (data not shown). Cell-fusion studies revealed that the mutation in Z65 was different from those previously noted for two complementation groups of fibroblasts from patients with Zellweger syndrome^{1,7} (unpublished results). We conclude that PAF-1 is specific for CHO cell mutant Z65 and is not primarily defective in these two types of fibroblasts. Northern blot analysis showed that size of the rat liver PAF-1 mRNA was about 1.9 kb, and the transcript of PAF-1 in Z65 was nearly the same both in size and amount as that in the wild-type cells (data not shown), thereby implying that a point mutation of PAF-1 is more likely to be the primary defect in the Z65.

Intracellular localization of PAF-1 was studied with an antibody against a synthetic PAF-1 peptide (Fig. 1b, dashed line). The antibody specifically immunoprecipitated the *in vitro* synthesized PAF-1 (Fig. 3a). PAF-1 showed unique property, the shift of mobility in SDS-PAGE before and after treatment with iodoacetamide (IAA) (Fig. 3a and b). Subcellular fractionation and immunoblots strongly suggested that the 35K PAF-1 protein was concentrated in the light mitochondrial fraction of

rat liver and was localized in peroxisomes, with evidence of concomitant localization of PAF-1 with that of peroxisomal proteins such as catalase and 22K integral membrane protein¹³ (Fig. 3c). PAF-1 behaves as an integral membrane protein of peroxisomes, assessed by sonication in salt, extraction with sodium carbonate¹⁴, and treatment with Triton X-114¹⁵ (Fig. 3d). These results are consistent with the presence of two possible membrane-spanning segments¹⁶ (Fig. 1b, underlined). Furthermore, PAF-1 was highly sensitive to externally added

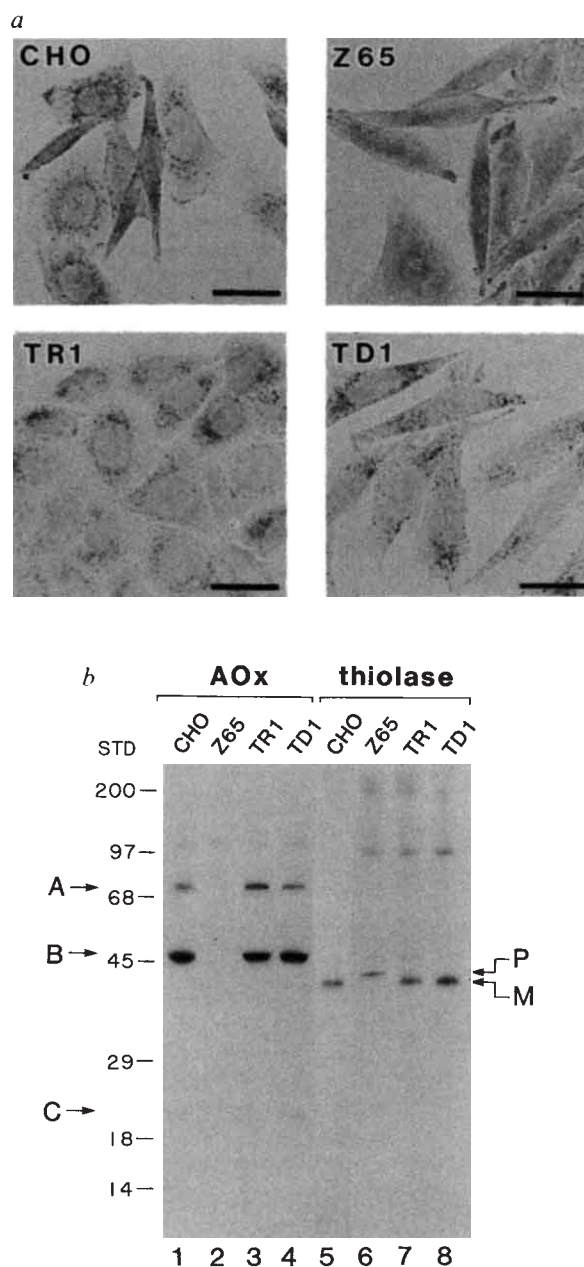


FIG. 2 Properties of PAF-1 transfected cells. *a*, Morphological analysis by immunocytochemistry. CHO, wild-type CHO cells; Z65, peroxisome-deficient Z65; TR1, primary cDNA library-transfectants of Z65 which were then selected by the P12/UV method (Table 1a); TD1, Z65 transfected with pcD65D. Bar, 15 μ m. *b*, Biosynthesis of peroxisomal enzymes. A, B and C indicate the positions of A, B and C components of acyl-CoA oxidase (AOx), respectively; P and M, a larger precursor and mature form of 3-ketoacyl-CoA thiolase, respectively. Molecular mass standards (STD) are on the left. METHODS. *a*, Immunocytochemical staining using anti-rat catalase antibody was performed as described⁶. *b*, Cells were labelled with [³⁵S]methionine for 42 h; AOx and thiolase were immunoprecipitated by specific antibodies and analysed by SDS-PAGE/fluorography⁶.

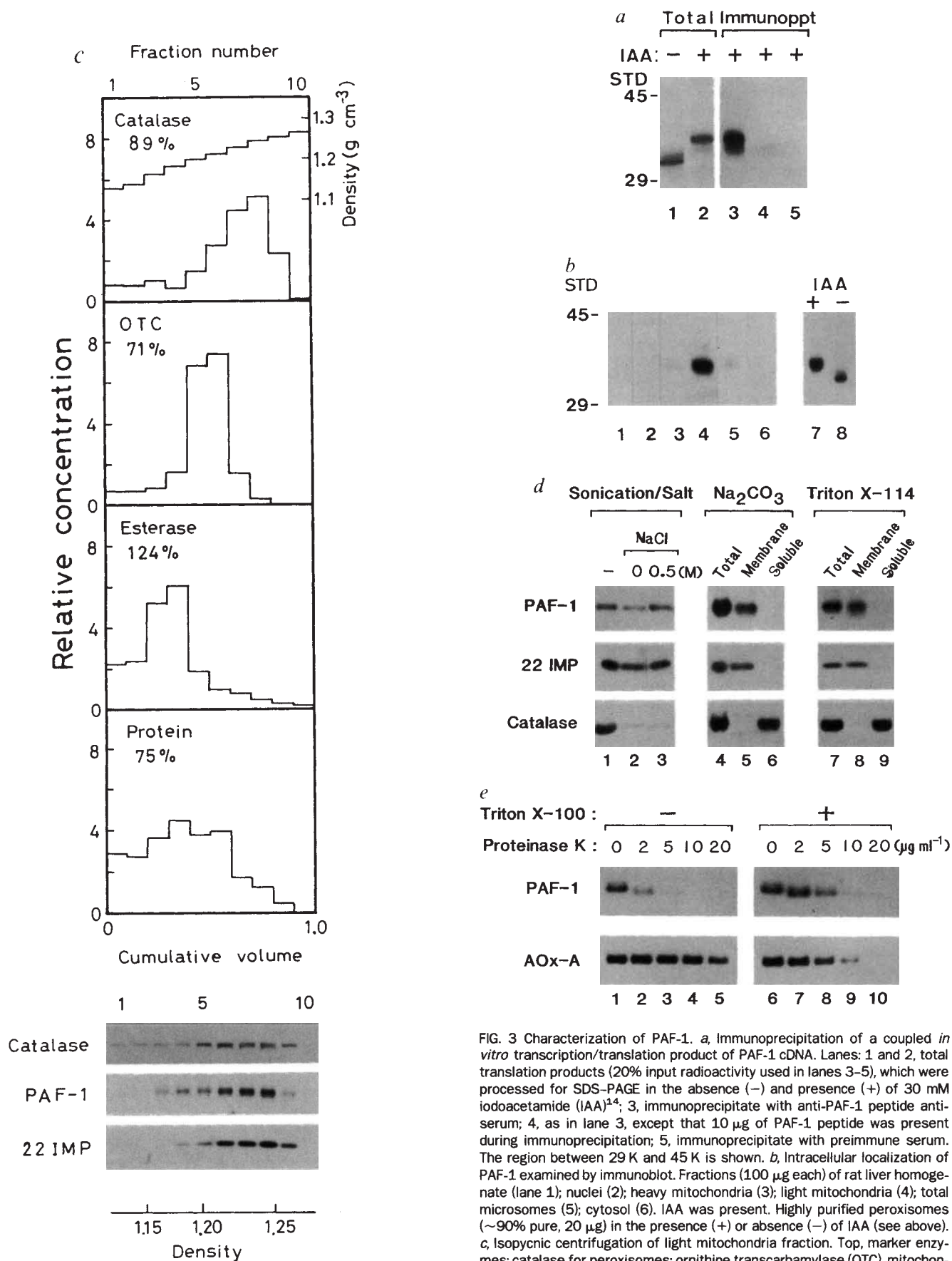


FIG. 3 Characterization of PAF-1. *a*, Immunoprecipitation of a coupled *in vitro* transcription/translation product of PAF-1 cDNA. Lanes: 1 and 2, total translation products (20% input radioactivity used in lanes 3–5), which were processed for SDS-PAGE in the absence (–) and presence (+) of 30 mM iodoacetamide (IAA)¹⁴; 3, immunoprecipitate with anti-PAF-1 peptide antiserum; 4, as in lane 3, except that 10 μg of PAF-1 peptide was present during immunoprecipitation; 5, immunoprecipitate with preimmune serum. The region between 29 K and 45 K is shown. *b*, Intracellular localization of PAF-1 examined by immunoblot. Fractions (100 μg each) of rat liver homogenate (lane 1); nuclei (2); heavy mitochondria (3); light mitochondria (4); total microsomes (5); cytosol (6). IAA was present. Highly purified peroxisomes (~90% pure, 20 μg) in the presence (+) or absence (–) of IAA (see above). *c*, Isopycnic centrifugation of light mitochondria fraction. Top, marker enzymes: catalase for peroxisomes; ornithine transcarbamylase (OTC), mitochondria; esterase, microsomes. Percentages, the recovery of each constituent. Bottom, immunoblot with antisera against catalase, PAF-1, and 22 K peroxisomal integral membrane protein (22 IMP)²⁵. *d*, Intraperoxisomal localization of PAF-1. Three different methods were used, sonication and salt wash

(lanes 1–3), treatments with sodium carbonate (4–6) and Triton X-114 (7–9). Immunoblot was done as in Fig. 3c. Lanes: 1, no treatment (–), 2 and 3, pellets after sonication in NaCl as indicated; 4 and 7, total peroxisomes; 5 and 8, membrane fractions; 6 and 9, soluble fractions. *e*, Protease-treatment of peroxisomes. Immunoblot with anti-PAF-1 and anti-AOx antisera; only the A component of AOx is shown, as the B component was found resistant to proteinase K, under the conditions used.

METHODS. *a*, 1.3 kb *Pst*I–*Bam*HI cDNA fragment of pcD65K was subcloned into Bluescript SK[–] (Stratagene). Anti-PAF-1 antibody was raised against the synthetic peptide (see Fig. 1b) conjugated to keyhole limpet haemocyanin²⁶. *A* coupled *in vitro* transcription/translation⁵ and immunoprecipitation⁶ were carried out as described. *b*, Subcellular fractions of the liver of a male Sprague–Dawley rat were prepared as described^{27,28}, analysed by SDS–PAGE, and transferred to nitrocellulose⁶. Immunoblot was done with the anti-PAF-1 antiserum; antibody was detected by [¹²⁵I]protein A and autoradiography⁶. Exposure: lanes 1–6, 2 days; lanes 7 and 8, 1 day. *c*, Light mitochondrial fraction (9 mg protein) was centrifuged into a linear gradient of 30–60% sucrose at 50,000 r.p.m. for 1.5 h in a Beckman VTi 65.2 rotor. The gradient was collected into 10 fractions which were assayed for density, protein, and marker enzymes specific for organelles^{14,28,29}. An equal volume (7 μ l, 5 μ g protein at fraction 8) was analysed by immunoblot. Exposure: catalase, 2 h; PAF-1, 4 days; 22 IMP, 2 days. *d*, Highly purified peroxisomes (10 μ g), in 50 mM HEPES–KOH, pH 7.6, 1 mM dithiothreitol (50 μ l), were sonicated for 2 min at 0 °C in the presence (0.5 M) or absence of NaCl, and then centrifuged at 100,000 $\times g$ for 30 min. The resulting pellets were analysed. Membranes and soluble fraction were prepared by the sodium carbonate method¹⁴ and by treatment with the detergent Triton X-114¹⁵, from 20 and 10 μ g peroxisomes, respectively. Exposure: PAF-1, 4 days except for 2 days (lanes 1–3); 22 IMP, 1 day; catalase, 3 h. *e*, Peroxisomes (10 μ g) were digested, in 50 mM HEPES–KOH, 0.25 M sucrose (30 μ l), on ice for 30 min with proteinase K at the concentrations indicated, in the absence (–) or presence (+) of Triton X-100. The reaction was terminated by addition of protease inhibitors⁵. Exposure: PAF-1, 4 days; AOx, 12 h.

proteinase K, where acyl-CoA oxidase, a matrix enzyme, was resistant but became accessible by addition of the detergent Triton X-100, thereby suggesting that PAF-1 is exposed on the outer surface of peroxisomes (Fig. 3e). PAF-1 does not contain at its C terminus the Ser-Lys-Leu-COOH sequence, a peroxisome-targeting signal^{4,5}. In addition, PAF-1 has seven cysteine residues in its C-terminal region and these could explain the mobility shift by IAA, as noted for other proteins such as metallothionein¹⁷. This cysteine-rich segment is similar to the zinc-binding domains of zinc-finger DNA-binding proteins¹⁸ but the significance is unknown.

Several possible functions of PAF-1 in peroxisome biogenesis can be speculated. PAF-1 may be a component of a putative signal receptor or of transmembrane import machinery of newly synthesized proteins, although the anti-PAF-1 C-terminal peptide antibody did not inhibit *in vitro* the import of acyl-CoA oxidase⁵ (data not shown). This part of PAF-1 was not essential for the complementing activity of PAF-1 (manuscript in preparation). The function of PAF-1 could be studied, as in the case of yeast mitochondrial outer-membrane components of the putative import machinery^{19,20}.

Our method is a direct cloning, by means of genetic complementation, of cDNA encoding a factor essential for assembly of organelles, taking advantage of a most effective P12/UV selection, rapid immunocytochemical staining of peroxisomes, and complete complementation assay. Peroxisome-deficient mutants of the yeast *Saccharomyces cerevisiae* have been isolated²¹. But as *S. cerevisiae* is evolutionally highly distant from humans, the CHO cell mutants and rat PAF-1 cDNA may be more useful to better understand human peroxisomal disorders. □

6. Tsukamoto, T., Yokota, S. & Fujiki, Y. *J. Cell Biol.* **110**, 651–660 (1990).
7. Brul, S. *et al. J. clin. Invest.* **81**, 1710–1715 (1988).
8. Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745–2752 (1987).
9. Zoeller, R. A., Morand, O. H. & Raetz, C. R. H. *J. Biol. Chem.* **263**, 11590–11596 (1988).
10. Breitman, M. L., Tsui, L.-C., Buchwald, M. & Siminovich, L. *Molec. cell. Biol.* **2**, 966–976 (1982).
11. Zoeller, R. A. *et al. J. Biol. Chem.* **264**, 21872–21878 (1989).
12. Miura, S. *et al. J. Biol. Chem.* **259**, 6397–6402 (1984).
13. Fujiki, Y., Fowler, S., Shio, H., Hubbard, A. L. & Lazarow, P. B. *J. Cell Biol.* **93**, 103–110 (1982).
14. Fujiki, Y., Hubbard, A. L., Fowler, S. & Lazarow, P. B. *J. Cell Biol.* **93**, 97–102 (1982).
15. Bordier, C. *J. Biol. Chem.* **256**, 1604–1607 (1981).
16. Kyte, J. & Doolittle, R. F. *J. mol. Biol.* **157**, 105–132 (1982).
17. Sone, T., Yamaoka, K., Minami, Y. & Tsunoo, H. *J. Biol. Chem.* **262**, 5878–5885 (1987).
18. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1–3 (1988).
19. Vestweber, D., Brunner, J., Baker, A. & Schatz, G. *Nature* **341**, 205–209 (1989).
20. Söllner, T., Griffiths, G., Pfaller, R., Pfanner, N. & Neupert, W. *Cell* **59**, 1061–1070 (1989).
21. Erdmann, R., Veenhuis, M., Mertens, D. & Kunau, W.-H. *Proc. natn. Acad. Sci., U.S.A.* **86**, 5419–5423 (1989).
22. Mori, M., Morris, S. M. & Cohen, P. P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3719–3783 (1979).
23. Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
24. Hess, R., Staubli, W. & Riess, W. *Nature* **208**, 856–858 (1965).
25. Fujiki, Y., Rachubinski, R. A. & Lazarow, P. B. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7127–7131 (1984).
26. Lerner, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3403–3407 (1981).
27. Berthet, J. & de Duve, C. *Biochem. J.* **50**, 174–181 (1951).
28. Leighton, F. *et al. J. Cell Biol.* **37**, 482–513 (1968).
29. Marshall, M. & Cohen, P. P. *J. Biol. Chem.* **247**, 1641–1653 (1972).

ACKNOWLEDGEMENTS. We thank H. Okayama for the pL2 plasmid, H. Mori and A. Noda for advice on the molecular biology, S. Hashimoto for the COS-7 cells, H. Moser for fibroblasts from Zellweger patients, I. Arai-Kasuya for technical assistance and M. Ohara for helpful comments.

Conversion by retinoic acid of anterior cells into ZPA cells in the chick wing bud

N. Wanek*, D. M. Gardiner, K. Muneoka & S. V. Bryant†

Developmental Biology Center, University of California, Irvine, California 92717, USA

IN recent years there has been considerable interest in the role of retinoic acid (RA) in vertebrate-limb pattern formation. When RA is applied to the anterior of the chick wing bud, a mirror-image duplication of the limb pattern develops that is identical to the pattern resulting from grafts of posterior tissue (zone of polarizing activity, or ZPA)¹. It has been proposed that position along the anterior–posterior axis in the chick limb is specified by a gradient of a diffusible factor produced by the ZPA². The ZPA-mimicking action of RA has led to the hypothesis that exogenously applied RA acts by providing graded spatial information across the anterior–posterior limb axis^{3,4}. An alternative interpretation is that RA changes anterior cells into ZPA cells, which in turn provide the actual pattern-duplicating stimulus^{3–5}; there is already some preliminary evidence that this occurs⁵. A hybrid interpretation has also been suggested whereby ZPA cells are formed in response to RA exposure and then begin to release retinoids that act as graded spatial cues³. We have used a functional assay⁶ to test anterior chick wing-bud cells for ZPA activity after exposure to RA. The results of our studies indicate that the action of RA is to change anterior cells into ZPA cells. Further, our results indicate that it is unlikely that RA-treated anterior cells then begin producing RA in such a way as to provide a graded positional signal.

Micro-carrier beads were preloaded with RA and implanted into the anterior margin of wing buds (Fig. 1)³. The loading regime and dose of RA were standardized to yield pattern-duplicating activity equivalent to that reported previously for ZPA grafts and ZPA-mimicking RA bead implants^{2,3} (Table 1). At various times after implanting a bead, we removed a wedge of tissue adjacent to the bead and grafted it into the anterior of an untreated host wing. To ensure that we were not merely using

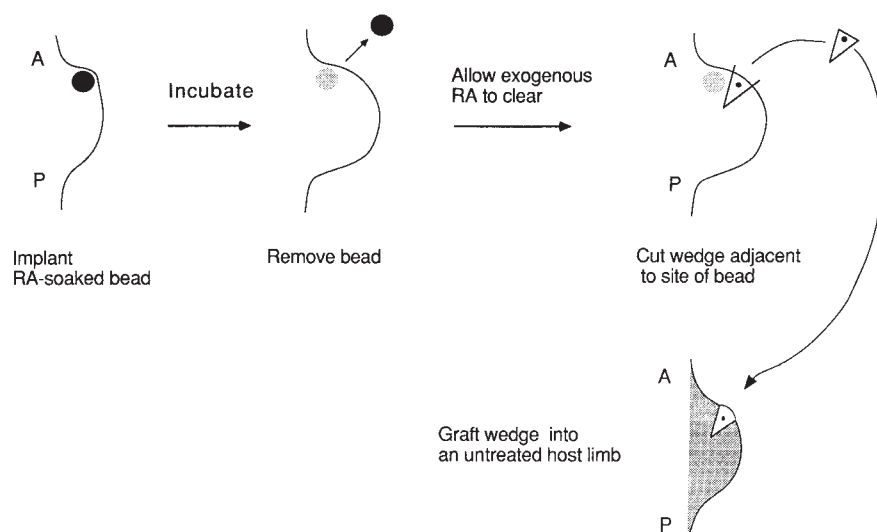
Received 24 September; accepted 13 December 1990.

1. Lazarow, P. B. & Moser, H. W. in *The Metabolic Basis of Inherited Disease* 6th edn (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 1479–1509 (McGraw-Hill, New York, 1989).
2. Verner, K. & Schatz, G. *Science* **241**, 1307–1313 (1988).
3. Gerace, L. & Burke, B. A. *Rev. Cell Biol.* **4**, 335–374 (1988).
4. Gould, S. J., Keller, G.-A. & Subramani, S. *J. Cell Biol.* **107**, 897–905 (1988).
5. Miyazawa, S. *et al. Molec. cell. Biol.* **9**, 83–91 (1989).

* Present addresses: Biology Department, Rochester Institute of Technology, Rochester, New York 14623, USA (N.W.); Biology Department, Tulane University, New Orleans, Louisiana 70118, USA (K.M.).

† To whom correspondence should be addressed.

FIG. 1 Functional assay for the conversion of anterior cells to ZPA cells. Six 200- μm -diameter Dowex AG1-X2 ion-exchange resin beads (Biorad) were soaked for 20 min at 20 °C in 100 μl of 1 mg ml⁻¹ solution of all-*trans* retinoic acid (Sigma) dissolved in DMSO. The chosen soaking dose is one that gives a range of duplications equivalent to standard ZPA grafts and to previously reported ZPA-mimicking RA-bead implants^{2,3}. Beads were rinsed in 100 μl Dulbecco's PBS (2 $\times$ 1 min) and stored in 3 ml PBS at 20 °C until single beads were implanted into the anterior margin of stage 20 chick wing buds¹⁹. 12–24 h after bead implantation, a carbon mark was placed in the limb bud immediately posterior to the bead. Beads were then removed and the embryos were incubated for an additional 90 min to clear any exogenous RA remaining in the tissue. A wedge (200–250 μm at the distal margin) was excised using tungsten knives and grafted into the anterior of an untreated stage 20 host wing bud. Embryos were checked for the presence of the graft after 24 h, and incubated further for 5 days. The anterior (A) and posterior (P) margins are marked.



the wedge to transfer exogenous RA to the host limb, we incubated the embryos for a minimum of 90 min between bead removal and excision of the wedges to allow the limb to clear RA remaining in the tissue (the half-life of exogenously applied RA in the chick wing bud is 20 min⁷). Wedges removed from control limbs and from limbs after 12 h of RA exposure did not induce any supernumerary digits in host limb-buds (Fig. 2a, and Table 2). Had the wedge grafts been transferring enough RA to induce supernumerary digits, we would have expected to see the strongest response at the 12-h time point as the tissue concentration of exogenous RA is higher at 12 hours than at later times⁷. Anterior wedges removed 15 h after implanting an RA-loaded bead did cause pattern duplications in 50% of the host limbs (Fig. 2b, and Table 2). The frequency of induced supernumerary digits rose to 92% after 24 h of RA treatment (Table 2). Thus, within 15 h of exposure to RA, anterior chick wing-bud cells have become functional ZPA cells as indicated by their ability to stimulate the formation of supernumerary digits in a host wing bud.

would contain the anlage of digits 4, 3 and 2. Instead, RA-treated tissue only contains cells that form the most extreme posterior margin of the limb. This result is expected if RA treatment converts anterior cells into ZPA cells.

Our results are consistent with the suggestion that the mode of action of exogenous RA involves two steps^{3,5,9}. RA initially converts anterior cells into ZPA cells, which then interact with adjacent anterior cells to bring about digit duplication. The dose-response to RA³, which mimics the response to variations in the number of grafted ZPA cells¹⁰, can be viewed as a consequence of increasing the numbers of anterior cells converted to ZPA cells with increasing exposure to RA³.

TABLE 1 Control RA-bead implants

Total <i>n</i>	Normal limbs	Limbs with supernumerary digits			
	2-3-4	Total	2-2-3-4	3-2-3-4	4-3-3-4
18	0 (0%)	18 (100%)	1	1	9

We have confirmed that the grafted cells induce new pattern in the host tissues (that is, they have ZPA activity), and that they are not simply undergoing self-differentiation in response to a gradient of RA originating from the implanted bead. To do this, we repeated our experiment using 24 h RA-treated anterior quail cells as the graft into anterior chick hosts. Analysis of the cellular contribution pattern revealed that the grafted, RA-treated anterior cells only contributed to the extreme posterior margin of the duplicated pattern, whereas the remainder of the induced pattern was formed by host chick cells (Fig. 2c). Control grafts of quail ZPA cells into anterior chick hosts yielded the same pattern of contribution (Fig. 2c).

Although the cumulative evidence suggests some role for retinoids in limb patterning, our results do not support the idea that a gradient of RA across the anterior–posterior axis sequentially specifies digits according to local, threshold RA concentrations^{3,4}. Based on the known range of action of the ZPA⁸, this interpretation predicts that the tissue adjacent to the RA-bead

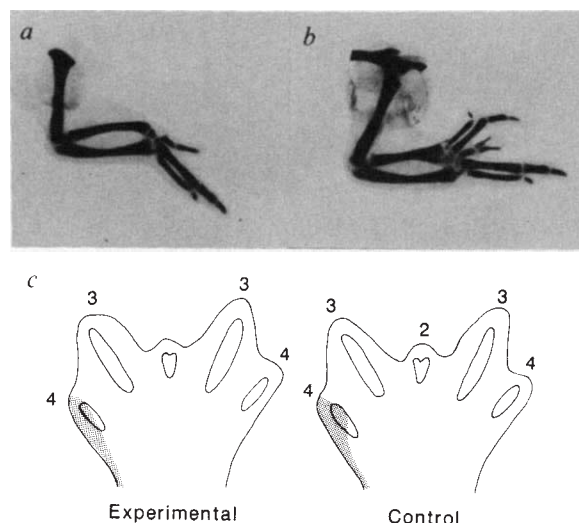


FIG. 2 Whole-mount skeletal preparations of limbs and patterns of cellular contribution from graft and host cells. *a*, Host limb that received an anterior wedge graft exposed to RA for 12 h showing normal skeletal pattern with digits 2,3,4. *b*, Host limb that received a wedge graft exposed to RA for 18 h. The autopod has six digits identified as 4-3-2-2-3-4. *c*, Contribution of graft (quail) and host (chick) cells in Feulgen-stained paraffin sections. Experimental grafts were wedges of anterior tissue from quail limb buds that had an RA-bead implanted into the anterior margin 24 h before grafting (see Fig. 1 for details). As a control, wedges of quail ZPA were grafted into the anterior margin of host chick wing buds. Limbs resulting from three control grafts and four experimental grafts all had duplicated digits and were analysed for cellular contribution. The data presented in this figure are for the experimental and control limbs that exhibited the maximum degree of contribution from the grafted quail tissue.

TABLE 2 Conversion of anterior cells to ZPA cells

Length of RA exposure (h)	n	Normal limbs 2-3-4	Total	Limbs with supernumerary digits				
				?-2-3-4 or ??-2-3-4	2-2-3-4	3-2-3-4 or 3-2-2-3-4	4-?-2-3-4 or ?-3-2-3-4	4-3-22-3-4 or 4-3-2-2-3-4
0*	15	15 (100%)	—	—	—	—	—	—
12	8	8 (100%)	—	—	—	—	—	—
15	8	4 (50%)	4 (50%)	2	—	—	2	—
18	11	4 (36%)	7 (64%)	1	3	2	—	1
24	13	1 (8%)	12 (92%)	1	4	3	1	3

* Control wedges from limbs implanted with beads soaked in dimethylsulphoxide only; wedges were removed and grafted either 18 ($n=6$) or 24 ($n=9$) hours after bead implantation.

This model leaves open the questions of how the placement of ZPA cells (either grafted or RA-induced) adjacent to anterior cells leads to digit duplication, and of the role of the ZPA in normal limb outgrowth. It has been suggested that the new ZPA could exert its effect by becoming the source of a new RA gradient^{3,9}. The duplicated digits would then be specified by the different levels of RA in the new gradient. Our results argue against a role for an RA gradient in digit specification for two reasons. First, as a gradient of exogenous RA did not initially specify any part of the pattern other than the ZPA, there is no reason to expect that a hypothetical, secondarily established RA gradient would do so. Secondly, if ZPA cells produce RA, it follows from our results that anterior cells must respond to RA exposure by producing RA, as they change into functioning ZPA cells. Hence, RA would be acting autocatalytically in the limb, stimulating cells to produce more RA. This behaviour is incompatible with the maintenance of a gradient of RA (either endogenous or exogenous) unless further suppositions are made.

One possible way for an autocatalytic molecule to be distributed non-uniformly is if an inhibitor exists to localize its distribution, as in a reaction-diffusion system^{11,12}. In the case of RA, a hypothetical inhibitor could confine RA to the posterior border of the limb bud where it has been reported to be more abundant⁴. But in such a model, RA (the activator) would establish the posterior margin of the limb-bud rather than provide a graded positional signal. The inhibitor would be graded across the limb-bud, and might therefore serve as a positional signal. There is no evidence at present for the existence of such a molecule in limbs.

An alternative interpretation of the role of exogenous RA in chick digit duplication is that after RA converts anterior cells to ZPA/posterior boundary cells, adjacent anterior and posterior cells interact directly, at close range, stimulating new growth and the formation of new pattern^{13,14}. The interpretation is fully consistent with the results reported here, with previous experimental findings in chick limbs^{14,15}, and with what is known about retinoids and their mode of action in this and other systems. It also encompasses the results of RA treatment on amphibian limbs, where RA has been shown to have similar dramatic effects on pattern formation in all three limb axes¹⁶⁻¹⁸. □

Received 29 October; accepted 21 December 1990.

1. Tickle, C., Alberts, B., Wolpert, L. & Lee, J. *Nature* **296**, 564-566 (1982).
2. Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199-202 (1975).
3. Tickle, C., Lee, J. & Eichele, G. *Dev. Biol.* **109**, 82-95 (1985).
4. Thaller, C. & Eichele, G. *Nature* **327**, 625-628 (1987).
5. Summerbell, D. & Harvey, F. *Prog. Clin. Biol. Res.* **110**, 109-118 (1983).
6. Smith, J. C. *J. Embryol. exp. Morph.* **52**, 105-113 (1979).
7. Eichele, G., Tickle, C. & Alberts, B. M. *J. Cell Biol.* **101**, 1913-1920 (1985).
8. Honig, L. *Nature* **291**, 72-73 (1981).
9. Smith, S. M. *et al. Development* **107** (suppl.) 121-131 (1989).
10. Tickle, C. *Nature* **289**, 295-298 (1981).
11. Gierer, A. & Meinhardt, H. *Kybernetik* **12**, 30-39 (1972).
12. Meinhardt, H. *Development* **107** (suppl.) 169-180 (1989).
13. French, V., Bryant, P. J. & Bryant, S. V. *Science* **193**, 969-981 (1976).
14. Bryant, S. V. & Muneoka, K. *Trends Genet.* **2**, 153-159 (1986).
15. Javois, L. C. in *Pattern Formation: A Primer in Developmental Biology* (eds Malacinski, G. M. & Bryant, S. V.) 557-579 (Macmillan, New York, 1984).

16. Maden, M. *Nature* **295**, 672-675 (1982).
17. Stocum, D. L. & Thoms, S. D. *J. exp. Zool.* **232**, 207-215 (1984).
18. Ludolf, D. C., Cameron, J. A. & Stocum, D. L. *Dev. Biol.* **140**, 41-52 (1990).
19. Hamburger, V. & Hamilton, H. L. *J. Morph.* **88**, 49-92 (1951).

ACKNOWLEDGEMENTS. We thank T. Hayamizu, A. Groell and L. Mullen for discussions and criticisms of the manuscript. This work was supported by a grant from the National Institutes of Health and by a gift from the Monsanto Company.

Retinoic acid induces polarizing activity but is unlikely to be a morphogen in the chick limb bud

Sumihare Noji*, Tsutomu Nohno†, Eiki Koyama‡, Kohki Muto‡, Kazuhiko Ohshima‡, Yoshinobu Aoki§, Koji Tamura§, Kojune Ohsugi§, Hiroyuki Ide§, Shigehiko Taniguchi* & Taiichi Saïto†

* Department of Biochemistry, ‡ Department of Oral and Maxillofacial Surgery I and II, Okayama University Dental School, 2-5-1 Shikata-cho, Okayama City 700, Japan

† Department of Pharmacology, Kawasaki Medical School, 577 Matsushima, Kurashiki City 701-01, Japan

§ Biological Institute, Tohoku University, Aoba, Sendai City 980, Japan

RETINOIC acid is a putative morphogen in limb formation in the chick and other vertebrates¹⁻⁵. In chick limb formation, it is thought that retinoic acid is released from the zone of polarizing activity (ZPA) and the concentration gradient of retinoic acid formed from the posterior to the anterior provides positional cues for digit formation^{1-4,6}. Implantation of a bead containing retinoic acid at the anterior margin of the limb bud induces a mirror-image symmetrical duplication of the digit pattern similar to that observed when the ZPA is grafted into the anterior margin of the host limb bud^{7,8}. Also, the level of endogenous retinoic acid (25 nM on average) is higher in the posterior one third of the limb bud^{1,6}. We found that when the bead containing either retinoic acid or an analogue but not the ZPA, was implanted in the anterior margin of the chick limb bud, expression of the retinoic acid receptor type- β gene was induced around the bead within 4 h. These results indicate that exogenous retinoic acid is not identical with the ZPA morphogen. As the anterior tissue exposed to retinoic acid has polarizing activity⁹, we conclude that the primary function of exogenous retinoic acid is to induce polarizing activity in the limb bud.

THE nucleotide and deduced amino-acid sequence of a chick retinoic acid receptor (RAR) complementary DNA are shown in Fig. 1. The amino-acid sequence is 95% homologous to the sequences of the mouse¹⁰ and human¹¹ RAR β proteins, but <75% homologous to those of the mouse RAR α and γ protein¹⁰. Thus, we concluded that the cloned chick RAR cDNA is of the β -subtype. It has been reported that the RAR β gene is expressed

as two transcripts of ~3.1 and 3.5 kilobases in the developing chick limb bud¹. We performed *in situ* hybridization on serial parasagittal and transverse sections of the limbs of stage 20–30 (Hamburger–Hamilton¹²) embryos with both sense and antisense riboprobes for the RAR β transcripts. Dense accumulation of white silver grains was observed (Fig. 2) under dark-field view in the proximal region of the limb buds and motoneurons of the spinal cord from 19–30 stages. Because no significant accumulation of grains was observed with the sense probe (data not shown), localization of grains with the antisense probe indicates the presence of transcripts of the β -subtype gene.

The RAR β gene was expressed in the proximal region of both wing and leg buds at stages 19–20 (Fig. 2*a* and *b*, respectively), 23 (Fig. 2*c*) and 26 (Fig. 2*d*), but not in the distal region, indicating no specific expression of the RAR β gene around the ZPA. At stages 26–30, transcripts were observed around the condensed mesenchyme of the developing cartilage of the limbs (Fig. 2*e*).

Transcription of the β gene is upregulated by retinoic acid in hepatoma cells¹³ and F9 EC cells¹⁰. Furthermore, the responsive element of the human and mouse β genes was a cis-activating element responsive to both RAR α and RAR β (refs 14, 15). Thus, when the bead containing retinoic acid or its analogue TTNPB (E)-4-[2-(5,6,7,8-tetrahydro-5,5,8,8-tetramethyl-2-naphthalenyl)-1-propenyl]benzoic acid was implanted in the anterior margin of the right limb bud (Fig. 3*a*) to induce duplication of the digits^{16,17}, we expected enhanced transcription of the β gene in the retinoic acid-treated right bud.

To test our hypothesis, we hybridized sections of the treated limb buds *in situ*. The transcripts were increased in the right wing buds implanted with the bead containing retinoic acid (Fig. 4*a*) or TTNPB (Fig. 4*b*) at a morphogenetic concentration when the buds were fixed at 4 h (Fig. 4), 8 h and 12 h (data not shown) after implantation but were not in the left one. At 24 h, however, significant induction was no longer observed in the treated limb buds, partly because of the depletion of exogenous retinoic acid. A novel inducer of digit duplication 3,4-didehydroretinoic acid¹⁸, also induced expression of the RAR β gene, like retinoic acid (Fig. 4*c*). When as a control a bead soaked in the dimethyl sulphoxide solution free of retinoic acid was implanted in the same region, there was no induction (data not shown). However, transcripts of the RAR α gene were uniformly distributed over the limb buds (data not shown), as already found in mouse^{19,20}; no significant increase in transcripts was observed around the TTNPB-treated bead in the limb buds (data not shown).

Even when the TTNPB-treated bead was implanted into the ZPA of the limb bud (Fig. 3*b*), β -gene expression in mesenchymal cells was induced around the bead (Fig. 4*d*). This shows that transcription of the β gene is inducible by exogenous retinoic acid in any mesenchymal cells of the limb bud at stages 20–21. Because no specifically enhanced expression of the β gene was observed around the ZPA of the normal limb bud (Fig. 2*a–c*), we assumed that sufficient retinoic acid is not released from the ZPA to increase the levels of transcripts of the β gene. To verify this, we performed *in situ* hybridization

	CCAACAGCCTACGTCCAAAAAGGGCAGAGTTTGATGGA	41
	GTTCGGGTGACTTTTATGTGCCATTTCTCCACACCTAGAGAAAGCACTTTTCAGGGCATCTGTACAGGGGAATC	120

A	ATG TTT GAC TGT ATG GAT GTT CTA GCA GTG AGC CCT GCT CAG ATG CTG GAT TTC TAC ACT Met Phe Asp Cys Met Asp Val Leu Ala Val Ser Pro Ala Gln Met Leu Asp Phe Tyr Thr	180 (20)
	GCG AGT CCG TCT TCC TGC ATG CTC CAG GAG AAG GCT CTC AAA GCA TGC TTC AGT GGA TTG Ala Ser Pro Ser Ser Cys Met Leu Gln Glu Lys Ala Leu Lys Ala Cys Phe Ser Gly Leu	240 (40)
	GCA CAA ACA GAG TGG CAA CAC CGG CAC AGT GCT CAA TCA GTT GAA ACT CAG AGC ACC AGT Ala Gln Thr Glu Trp Gln His Arg His Ser Ala Gln Ser Val Glu Thr Gln Ser Thr Ser	300 (60)
B	TCT GAG GAA CTT GTT CCA AGC CCT CCT TCA CCA CTT CCA CCC CCG CGT GTT TAC AAG CCC Ser Glu Glu Val Pro Ser Pro Pro Ser Pro Leu Pro Pro Pro Arg Val Tyr Lys Pro	360 (80)
	TGT TTT GTC TGT CAA GAC AAA TCA TGT GGA TAT CAC TAT GGT GTC AGT GCT TGT GAG GGA Cys Phe Val Cys Gln Asp Lys Ser Ser Gly Tyr His Tyr Gly Val Ser Ala Cys Glu Gly	420 (100)
	TGT AAG GGC TTT TTC CGC AGG AGT ATC CAG AAG AAC ATG GTT TAC ACA TGT CAT AGA GAT Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Lys Asn Met Val Tyr Thr Cys His Arg Asp	480 (120)
C	AAG AAC TGT GTT ATC AAT AAA GTT ACC AGA AAT GCG TGC CAG TAC TGG AGA CTA CAG AAG Lys Asn Cys Val Ile Asn Lys Val Thr Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys	540 (140)
	TGC TTT GAA GTG GGA ATG TCC AAA GAA TCT GTC AGA AAT GAT AGG AAC AAA AAG AAG AAG Cys Phe Glu Val Gly Met Ser Lys Glu Ser Val Arg Asn Asp Arg Asn Lys Lys Lys Lys	600 (160)
	GAA CCT ACA AAG CAG GAG TCC ACA GAA AAC TAT GAA ATG ACA GCA GAG CTG GAT CAT GTC Glu Pro Thr Lys Glu Gln GAG Met Thr Glu Asn Tyr Glu Met Thr Ala Glu Leu Asp Asp Leu	660 (180)
D	ACT GAG AAG ATC CGA AAA GCT CAC CAG GAA ACC TTC CCC TCG CTC TGC CAG CTG GGA AAG Thr Glu Lys Ile Arg Lys Ala His Gln Glu Thr Phe Pro Ser Leu Cys Gln Leu Gly Lys	720 (200)
	TAC ACT ACA AAC TCG AGT GCA GAC CAT CGG GTT CGA CTG GAC CTT GGG CTC TGG GAC AAA Tyr Thr Thr Asn Ser Ser Ala Asp His Arg Val Arg Leu Asp Leu Gly Leu Trp Asp Lys	780 (220)
	TTC AGT GAA CTT GCA ACA AAG TGC ATT ATT AAA ATA GTG GAA TTT GCA AAA CGA TTG CCA Phe Ser Glu Leu Ala Thr Lys Cys Ile Ile Lys Ile Val Glu Phe Ala Lys Arg Leu Pro	840 (240)
	GGC TTT ACA AGC TTG ACC ATT GCA GAC CAG ATA ACT CTA CTT AAA GCA GCG TGC CTG GAC Gly Phe Thr Ser Leu Thr Ile Ala Asp Gln Ile Thr Leu Leu Lys Ala Ala Cys Leu Asp	900 (260)
	ATA CTG ATC CTC AGA ATT TGC ACA AGA TAC ACA CCA GAA CAA GAC ACC ATG ACA TTT TCA Ile Leu Ile Leu Arg Ile Cys Thr Arg Tyr Thr Pro Glu Gln Asp Thr Met Thr Phe Ser	960 (280)
	GAT GGC CTT ACC CTG AAC CGA ACT CAG ATG CAC AAT GCT GGA TTT GGC CCC CTG ACC GAC Asp Gly Leu Thr Leu Asn Arg Thr Gln Met His Asn Ala Gly Phe Gly Pro Leu Thr Asp	1020 (300)
E	CTC GTG TTC ACC TTT GGC AAC CAG CTC CTG CCT TTG GAA ATG GAT GAC ACA GAA ACC GGT Leu Val Phe Thr Phe Ala Asn Gln Leu Leu Pro Leu Glu Met Asp Asp Thr Glu Thr Gly	1080 (320)
	CTC CTT AGT GGC ATC TGC TTG ATC TGT GGA GAT CGC CAG GAC CTT GAA GAA CCA ATG AAA Leu Leu Ser Ala Ile Cys Lys Leu Ile Cys Gly Asp Arg Gln Asp Leu Glu Glu Pro Met Lys	1140 (340)
	GTG GAT AAG CTT CAG GAA CCG TTG CTT GAA GCG CTG AAA ATC TAT ATC CGA AAG AGA CGA Val Asp Lys Leu Gln Glu Pro Leu Leu Glu Ala Leu Lys Ile Tyr Ile Arg Lys Arg Arg	1200 (360)
	CCC AAC AAG CCT CAC ATG TTT CCA AAG ATC TTA ATG AAA ATC ACA GAT CTT CGT AGC ATC Pro Asn Lys Pro His Met Phe Pro Lys Ile Leu Met Lys Ile Thr Asp Leu Arg Ser Ile	1260 (380)
	GAT GCA AAA GGT GCA GAA CGT GTC ATT ACG TTG AAA ATG GAA ATT CCT GGA TCC ATG ACA Ser Ala Lys Gly Ala Glu Arg Val Ile Thr Leu Lys Met Glu Ile Pro Gly Ser Met Pro	1320 (400)
	CCT CTT ATT CAA GAA ATG TTG GAG AAC TCT GAA GGA CAC GAA CCA CTG ACA CCA ACC TCA Pro Leu Ile Gln Glu Met Leu Glu Asn Ser Glu Gly His Glu Pro Leu Thr Pro Thr Ser	1380 (420)
F	AAT GGA AAT ACT GCA GAA CAC AGT CCG AGT ATC TCA CCG AGC TCA GTG GAT AAC AGC AGT Asn Gly Asn Thr Ala Glu His Ser Pro Ser Ile Ser Pro Ser Val Asp Asn Ser Ser	1440 (440)
	GTC AGT CAA TCC CCT ATG GTG CAA TAA GACATTTTCAGCTACTTCAGACATTCCTAACCTTTATGTAC Val Ser Gln Ser Pro Met Val Gln ***	1510 (448)
	AGGGATGAAAGCAAGAAAAACATTTTACTGCTGCTTACTTCTGGACTTAATATATATCTATATATATATATA	1589
	TATATAGATAAGAACTCAACAGGACCAAGAAATTTTCGTATGATCAATATATATCTCTGCTGTTAACTCTTAAA	1668
	ATAGAAAATGCAAACTTTTGAATCTT	1695

FIG. 1 The nucleotide and deduced amino-acid sequence of chick RAR cDNA. Domains A–F are indicated on the left side of the sequence. *** In frame stop codon. EMBL/GenBank/DBJ accession number: X57340.

METHODS. Chick embryos (White Leghorn) were purchased from Fukuda Co. (Okayama, Japan). Messenger RNAs from one hundred 4-day embryos were used to construct a cDNA library in a λ gt10 vector. The library (about 10^6 recombinant phage) was screened with a probe obtained by polymerase chain reaction (PCR) amplification with primers derived from conserved nucleotide sequences of human¹¹ and mouse RAR β genes¹⁰. One of ten clones isolated was subcloned into an M13 vector and sequenced by the dideoxy-chain termination method²⁴. A cDNA fragment was subcloned into a pGEM7 vector to synthesize riboprobes for *in situ* hybridization.

FIG. 2 Expression patterns of the chick *RAR β* gene, as revealed by *in situ* hybridization on parasagittal sections of the wing bud at stage 19(a), leg bud at stage 20(b), wing bud at stage 23(c), and transverse sections of the developing wing at stage 26(d) and 28 (arrow indicates the condensed mesenchyme) (e). Bright-(left) and dark-field (right) views. A, anterior; P, posterior; D, dorsal; V, ventral. Bar, 500 μ m.

METHODS. Chick embryos at varied stages were fixed with 4% paraformaldehyde, dehydrated with ethanol and then embedded in paraffin. Procedures for *in situ* hybridization were described previously²⁵⁻²⁷. Under our stringent washing conditions, there was no significant cross hybridization to transcripts of other subtypes on sections of mouse embryos, even though whole cDNAs for mouse RARs were used as templates for synthesis of riboprobes²⁷. Thus, we used the whole cDNA without a poly(A) tail as a template to synthesize riboprobes specific for the β subtype.

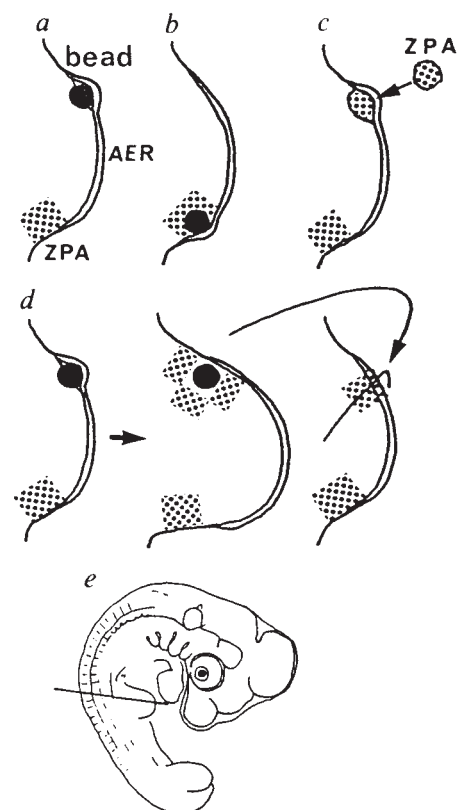
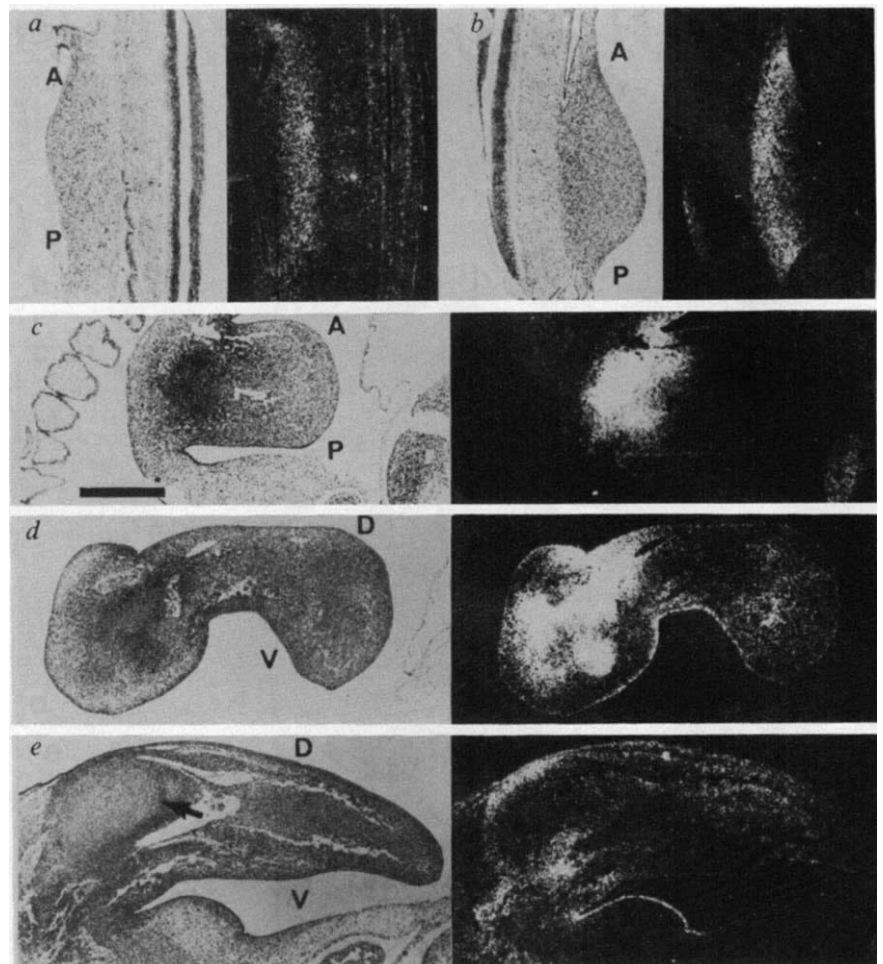


FIG. 3 Schematic illustrations of bead implantation into the anterior (a) and posterior margins (around the ZPA) (b) of the right wing bud, and the ZPA grafting to the anterior margin of the host right wing bud (c). d, Procedures for grafting tissues previously exposed to retinoic acid: a bead soaked in 0.1 mg ml⁻¹ retinoic acid was implanted in the anterior margin of the limb bud at stage 20. A tissue block around the bead was excised 12-15 h after implantation and then grafted in the anterior margin of the host right wing bud. Duplicated digits emerged after 7 days. e, Lateral aspect of the chick embryo at stage 22. The transverse sections of the limb buds were obtained by sectioning in the direction along the line indicated. ZPA, zone of polarizing activity; AER, apical ectodermal ridge; bead, a resin bead soaked in retinoic acid or TTNPB.

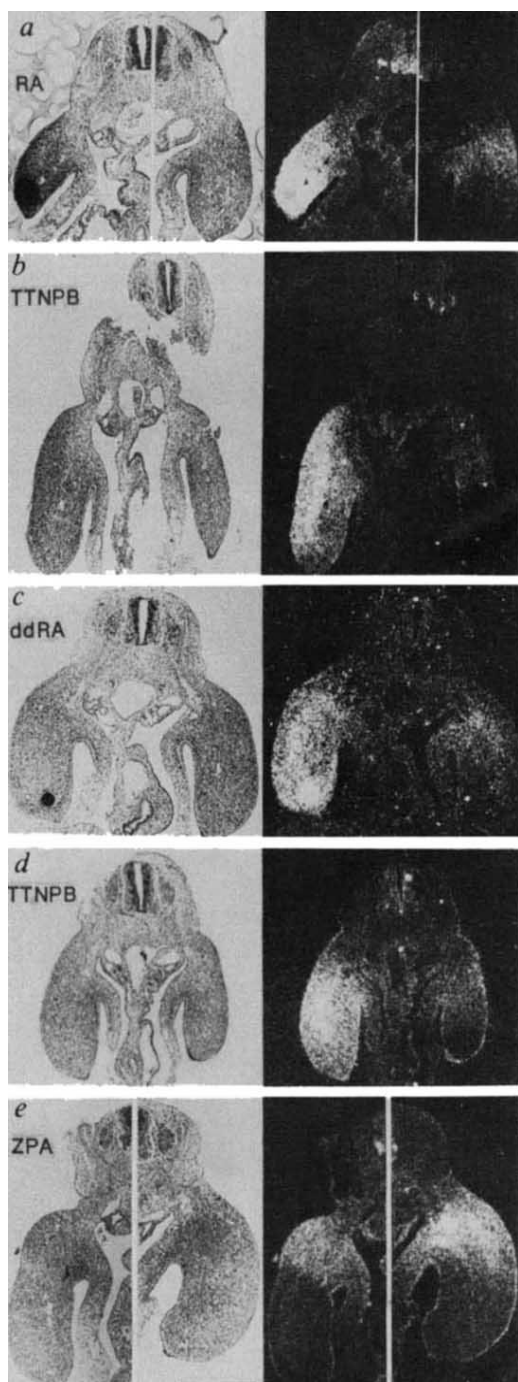


FIG. 4 Induction of chick *RARβ* gene expression by exogenous retinoic acid (RA), and RA analogue (TTNPB), or 3,4-didehydroretinoic acid (ddRA) released from the bead implanted in the anterior (a–c) and posterior (d) margins of the right wing bud at stage 20. *In situ* hybridization on transverse sections of the right wing buds, fixed at 4 h (a–d) after bead implantation. e, Expression of the *RARβ* gene in the ZPA-grafted limb bud (typical of six independent samples) fixed at 12 h after grafting in the anterior margin of the right limb bud. No significant induction was observed in the corresponding left wing bud of the same embryo as a control. 'RA', 'TTNPB', 'ddRA' and 'ZPA' indicate implantation of a bead containing RA, TTNPB, ddRA, and grafting of the ZPA, respectively. Bar, 500 μ m.

METHODS. Beads (AG1-X2, BioRad, in formate form, 200 μ m diameter) were soaked for 20 min at room temperature in dimethyl sulphoxide solution containing morphogenetic concentration of retinoids: 0.1 mg ml⁻¹ RA, 0.1 mg ml⁻¹ ddRA or 14 μ g ml⁻¹ TTNPB (Hoffmann-La Roche). Implantation of the bead into the limb bud was according to refs 21 and 28. Duplication of the digits was observed at 4–6 days after implantation. Several treated embryos were fixed with 4% paraformaldehyde at 4, 8, 12, 24, 48 and 72 h after implantation. The ZPA was grafted to the anterior margin of the right wing bud according to the method of G. Eichele (personal communication) (Fig. 3c) or Summerbell²⁹. Reproducibility of digit duplication by the ZPA-grafting is more than 70% and 90% by the Eichele and Summerbell methods, respectively.

containing a high concentration (16 mg ml⁻¹). We have used a well characterized bead method²² to confirm that this was not the case (Fig. 3d). The reproducibility of our observation on digit duplication, depending on the practical grafting conditions, was more than 60% when the tissues exposed to retinoic acid for 12–15 h were grafted to the anterior margin of the host limb bud (Fig. 3d). The positional information hypothesis^{1,3} postulates that digit pattern depends on the retinoic acid concentration, but this does not necessarily mean that the concentration gradient is a positional cue. Alternatively, the number of cells converted to the ZPA may depend on the retinoic acid concentration, because the digit pattern depends on the cell number of the ZPA (ref. 23).

Although our findings cannot be interpreted in terms of a simple model in which retinoic acid is important for positional information, we cannot rule out 'endogenous' retinoic acid as a morphogen. Auxiliary substances that suppress expression of the β gene in the ZPA may exist, or the responsiveness of mesenchymal cells to endogenous retinoic acid may be different from exogenous retinoic acid. If so, ZPA cells induced by exogenous retinoic acid may release 'endogenous' retinoic acid. □

Received 25 September 1990; accepted 22 January 1991.

- Eichele, G. *Trends Genet.* **5**, 246–251 (1989).
- Brookes, J. P. *Neuron* **2**, 1285–1294 (1989).
- Wolpert, L. *Development* **107** (suppl.), 3–12 (1989).
- Summerbell, D. & Maden, M. *Trends Neurosci.* **13**, 142–147 (1990).
- Ide, H. *Dev. Growth Different.* **32**, 1–8 (1990).
- Thaller, C. & Eichele, G. *Nature* **327**, 625–628 (1987).
- Tickle, C., Alberts, B. M., Lee, J. & Wolpert, L. *Nature* **296**, 564–565 (1982).
- Summerbell, D., *J. Embryol. exp. Morph.* **78**, 269–289 (1983).
- Summerbell, D. & Harvey, F. *Prog. Clin. Biol. Res.* **110**, 109–118 (1983).
- Zelen, A., Krust, A., Petkovich, M., Kastner, P. & Chambon, P. *Nature* **339**, 714–717 (1989).
- de Thé, H., Marchio, A., Tiollais, P. & Dejean, A. *Nature* **330**, 667–670 (1987).
- Hamburger, V. & Hamilton, H. *J. Morph.* **88**, 49–92 (1951).
- de Thé, H., Marchio, A., Tiollais, P. & Dejean, A. *EMBO J.* **8**, 429–433 (1989).
- de Thé, H., Vivanco-Ruiz, M., Tiollais, P., Stunnenberg, H. & Dejean, A. *Nature* **343**, 177–180 (1990).
- Sucov, H. M., Murakami, K. K. & Evans, R. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5392–5396 (1990).
- Eichele, G. & Thaller, C. *J. Cell Biol.* **105**, 1917–1923 (1987).
- Tickle, C., Lee, J. & Eichele, G. *Dev. Biol.* **109**, 82–95 (1985).
- Thaller, C. & Eichele, G. *Nature* **345**, 815–819 (1990).
- Noji, S. *et al. FEBS Lett.* **259**, 86–90 (1989).
- Dollé, P. *et al. Nature* **342**, 702–705 (1989).
- Oliver, G., Robertis, E. M., Wolpert, L. & Tickle, C. *EMBO J.* **9**, 3093–3099 (1990).
- Eichele, G., Tickle, C. & Alberts, B. M. *Analyt. Biochem.* **142**, 542–555 (1984).
- Tickle, C. *Nature* **289**, 295–298 (1981).
- Sanger, F., Nicklen, S., & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5468 (1977).
- Noji, S. *et al. Acta Histochem. Cytochem.* **23**, 353–366 (1990).
- Noji, S., Yamaai, T., Koyama, E., Nohno, T. & Taniguchi, S. *FEBS Lett.* **257**, 93–96 (1989).
- Osumi-Yamashita, N. *et al. FEBS Lett.* **264**, 71–74 (1990).
- Tamura, K., Ohsugi, K. & Ide, H. *Cell Dif. Dev.* **32**, 17–26 (1990).
- Summerbell, D. *J. Embryol. exp. Morph.* **32**, 227–237 (1974).

ACKNOWLEDGEMENTS. We thank Hoffmann-La Roche Co. for TTNPB and ddRA. This work was supported in part by the Japanese Ministry of Education, Science and Culture.

on sections of the limb bud grafted with the ZPA at the anterior margin (Fig. 3c). We observed no significant induction of *RARβ* gene expression around the grafted ZPA at 6 and 12 h after the grafting (Fig. 4e). Hence, it is unlikely that exogenous retinoic acid is identical with a morphogen secreted from the grafted ZPA. This is consistent with the finding that ectodermal expression of a homeobox gene *X1Hbox 1* in wing buds is enhanced by retinoic acid treatment but not by ZPA grafting²¹.

Our results suggest a two-step model for the role of exogenous retinoic acid in digit duplication, in which it first converts anterior mesenchymal cells to the ZPA, and then the ZPA cells secrete a morphogen. This model is consistent with the observation⁹ that a tissue previously exposed to retinoic acid, when grafted to the anterior margin of the other host limb bud, can induce additional digits, as observed for the ZPA-grafting. However, in these experiments the grafted tissue may have retained some retinoic acid from paper soaked in a solution

Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase

Nobutoshi Ito, Simon E. V. Phillips, Conrad Stevens, Zumrut B. Ogel, Michael J. McPherson, Jeffery N. Keen, Kapil D. S. Yadav & Peter F. Knowles

Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

GALACTOSE oxidase is an extracellular enzyme secreted by the fungus *Dactylium dendroides*. It is monomeric, with a relative molecular mass of 68,000, catalyses the stereospecific oxidation of a broad range of primary alcohol substrates and possesses a unique mononuclear copper site essential for catalysing a two-electron transfer reaction during the oxidation of primary alcohols to corresponding aldehydes¹. Recent evidence² arguing against a Cu(III)–Cu(I) couple³ implies the existence of a second redox-active site proposed to involve pyrroloquinoline quinone⁴ or a tyrosine radical^{5,6}. We now report the crystal structure of galactose oxidase at 1.7 Å resolution. This reveals a unique structural feature at the copper site with a novel thioether bond linking Cys 228 and Tyr 272 in a stacking interaction with Trp 290. We propose that these molecular components stabilize the protein free-radical species essential for catalysis and thus provide a 'built-in' secondary cofactor. This feature may represent a new mechanism for mediating electron transfer in metalloenzymes in the absence of exogenous cofactors.

Galactose oxidase (EC 1.1.3.9) was crystallized from 0.80 M acetate buffer (pH 4.3 to 4.7) with ammonium sulphate as precipitant. The space group is *C2* with unit cell parameters $a = 98.0$, $b = 89.4$, $c = 86.7$ Å, $\beta = 117.8^\circ$, and Table 1 summarizes the crystallographic data. The structure was solved by the multiple isomorphous replacement (m.i.r.) method using three heavy-atom derivatives. The initial map at 2.5 Å was of excellent quality, and most of the side chains as well as the backbone were easily located. The model was refined to 1.7 Å resolution to allow detailed structural analysis.

N-terminal amino-acid sequence data from galactose oxidase

and various protease V8- and cyanogen bromide-derived peptides were determined and used to design redundant primers for the polymerase chain reaction (PCR). A unique DNA fragment of 1.4 kilobase (kb) was amplified from *D. dendroides* genomic DNA and used as a homologous probe to isolate a galactose oxidase clone from a genomic library. The DNA sequence of the galactose oxidase gene (M.J.M., C.S., Z.B.O., P.F.K., J.K. & K.D.S.Y., manuscript in preparation) is consistent with all available peptide sequence data (~50% of the protein sequence) with the exception of Tyr 272 (see below). Translation of the DNA sequence would produce a mature galactose oxidase protein of 639 amino acid residues; the complete sequence has been fitted to the electron density map.

Galactose oxidase consists of three predominantly β -structure domains (Fig. 1a) with only a single α helix (residues 327–332). This preponderance of β structure both within and between domains probably contributes to the remarkable stability of the enzyme, which is active in 6 M urea⁷. The first domain (residues 1–155) has a β -sandwich structure and is linked to the second domain by a well-ordered stretch of polypeptide chain. The second and largest domain (residues 156–532) has pseudo-sevenfold symmetry with an overall appearance of a seven-petalled flower, where each petal consists of a four-stranded antiparallel β sheet (Fig. 1b). A similar arrangement has also been reported in methylamine dehydrogenase⁸ and neuraminidase⁹ although only six sheets, rather than seven, are evident in the latter case. The cupric ion resides on the solvent-accessible surface of this domain close to the pseudo-sevenfold axis. The third domain (residues 533–639) lies on the opposite side of the second domain from the copper. Two of the seven β strands in the third domain pierce the middle of the second domain along the pseudo-sevenfold axis and provide a ligand to the copper (Fig. 1a). The gap between this 'finger' and the second domain is filled with ~40 well-defined water molecules. All electron density observed can be accounted for by the known primary structure of the protein and by the solvent and ions in the solution; no density corresponding to pyrroloquinoline quinone (PQQ; ref. 4) or carbohydrate is evident.

Figure 2a shows the region around the cupric ion to be extremely rich in aromatic side chains. The copper site shown schematically in Fig. 2b has five ligands in square pyramidal coordination. Tyr 272, His 496, His 581 and an acetate ion form an almost perfect square with respective distances from the copper of 1.94, 2.11, 2.15 and 2.27 ± 0.15 Å. The fifth, axial,

TABLE 1 Statistics of data collection and refinement

Crystal	Resolution (Å)	N_{OBS}	N_{UNIQUE} (% of total possible)	R_{sym}^* (%)	$\Delta_{\text{iso}}^\dagger$ (%)	Number of sites	$R_{\text{centric}}^\ddagger$ (%)
Native (pH 4.5)	10–1.7	117,980	57,092 (79.1)	3.9			
	(1.8–1.7)	11,365	5,908 (52.6)§				
Native (pH 7.0)	10–1.9	83,387	46,302 (89.5)	5.0			
	(2.0–1.9)	7,378	5,589 (75.8)§				
K ₂ PtCl ₄	20–2.3	44,990	22,292 (75.8)	5.4	16.1	1	54
H ₂ IrCl ₆	20–2.2	54,428	27,817 (82.4)	4.4	14.9	3	45
Pb(NO ₃) ₂	20–2.2	54,932	27,703 (81.6)	4.9	13.9	10	56

Diffraction data from a native crystal were collected to 1.7 Å resolution on an area detector (Nicolet/Siemens X100A) mounted on a rotating anode X-ray generator (Rigaku RU-200), as well as from three heavy-atom derivatives (K₂PtCl₄, H₂IrCl₆, Pb(NO₃)₂). The frame data were processed with the program XDS (ref. 27), and all other crystallographic calculations were carried out with the CCP4 program suite (SERC Daresbury Laboratory, UK). A 2.5 Å m.i.r. map was calculated using both isomorphous and anomalous contributions with an average 'figure of merit' of 0.58. After ten cycles of solvent flattening²⁸, the polypeptide chain was fitted to the map using the program FRODO (ref. 29). The molecular model (including 316 water molecules, 2 acetate ions and a sodium ion) was refined by the Hendrickson–Konnert restrained least squares method to $R = 0.177$ for all the reflections between 10 and 1.7 Å. Root-mean-square deviation from ideal bond length is 0.018 Å, and all but three residues lie in the 'allowed' regions in the Ramachandran plot. Another native data set was collected to 1.9 Å from a crystal soaked in 0.1 M PIPES buffer at pH 7.0, and this structure (including 310 water molecules and a sodium ion) was refined to $R = 0.170$.

* $R_{\text{sym}} = \sum_i \sum_j |I(h_i) - I(h_j)| / \sum_i \sum_j I(h_i)$ where h are unique reflection indices.

† Fractional isomorphous difference (20–2.5 Å).

‡ $R_{\text{centric}} = (\sum_i |F_{\text{PH}} - F_P| - |F_H|) / \sum_i |F_{\text{PH}} - F_P|$ (20–3.0 Å).

§ These numbers are for the highest resolution shell.

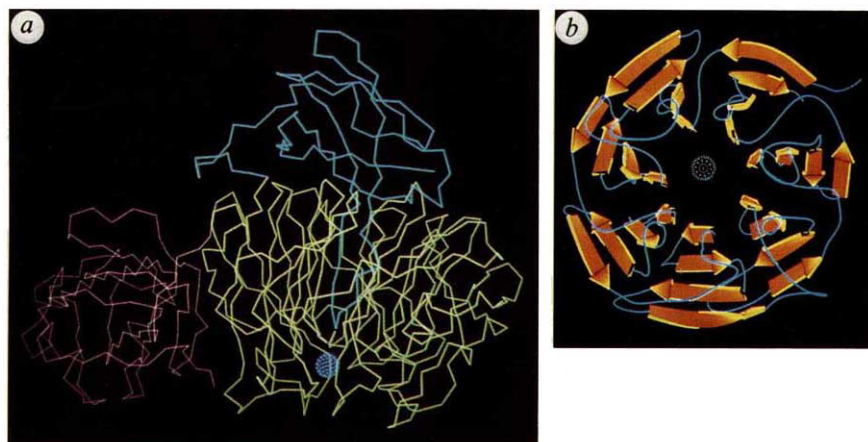


FIG. 1 *a*, C α backbone of galactose oxidase. The complete molecule is coloured according to domains (the first domain in red, the second yellow and the third blue). *b*, Computer-generated ribbon diagram³⁰ of the second domain, looking along the pseudo-sevenfold axis (the short α helix is not highlighted). It should be noted that there is a large hole along the axis, which is filled by two β strands from the third domain and a number of water molecules. The small blue sphere is the cupric ion.

ligand (Tyr 495) is 2.69 Å from the copper. In PIPES buffer (pH 7.0), the structure is essentially the same as in acetate buffer (pH 4.5) although the acetate ion ligand is replaced by a smaller molecule, probably water, at a distance of 2.8 Å, which is too long for coordination to the copper. The structure of the copper site at pH 7.0, where galactose oxidase is active, could be described as having distorted tetrahedral geometry, which would facilitate the redox change from Cu(II) to Cu(I) after addition of the alcohol substrate⁵. It is noteworthy that the structure of the active site in the hexachloroiridate heavy-atom derivative, which should be fully oxidized⁵, is indistinguishable from that of the native enzyme. Because the native enzyme is a mixture of activated and reversibly inactivated forms⁵, this implies that there is no major alteration in the atomic arrangement of the active site structure between these forms.

It is of particular interest that one of the copper ligands, Tyr 272, is covalently bound at C ϵ to the sulphur atom of Cys 228 (Fig. 3*a, b*). This bond seems to have partial double-bond character (as suggested by its *cis* conformation) with C β of Cys 228 lying in the same plane as the ring of Tyr 272, suggesting an extended aromatic system. The indole ring of Trp 290 is stacked on this 'pseudo side chain', with its six-membered ring located exactly above the sulphur atom (the distances between the sulphur and these six carbons of Trp 290 are 3.84 ± 0.04 Å) (Fig. 3*c*). This model strongly suggests the involvement of a free radical associated with this tyrosine as the secondary cofactor rather than PQQ (refs 3–5). The unusual bond between Cys 228 and Tyr 272 probably plays an essential part in the catalytic mechanism. The sulphur might assist delocalization of the free

radical with the stacking tryptophan aiding its stabilization. The presence of the 'Tyr-Cys bridge' identified from the electron-density map is supported by several experimental facts. The peptide sequencing has repeatedly failed to identify Tyr 272 in a peptide obtained by endoproteinase Lys-C digestion of carboxymethylated galactose oxidase (M.J.M., C.S., Z.B.O., P.F.K., J.K. & K.D.S.Y., manuscript in preparation). Resonance Raman data for galactose oxidase shows systematically and significantly lower frequencies for all the ring modes of a copper-bound tyrosine than found previously in other metalloenzymes⁶. Titration of the reduced and unfolded enzyme by 5,5'-dithio-bis(2-nitrobenzoic acid) indicates the presence of only five cysteines⁷ whereas our DNA sequence shows six. Furthermore, amino-acid analysis of acid-hydrolysed galactose oxidase reveals an unusual peak in the aromatic region¹⁰. Further experiments to confirm this novel bridge-structure by chemical methods are in progress. A similar thioether bond between a cysteine and a histidine has been identified by chemical analysis in another copper-containing enzyme, tyrosinase, from *Neurospora crassa*¹¹ which suggests that such a structural feature may occur more widely. It is tempting to speculate that certain metalloenzymes have evolved a mechanism for circumventing the requirement for exogenous secondary cofactors by developing a 'built-in' cofactor based, in the case of galactose oxidase, on a stabilized free radical.

Our model suggests two possibilities for substrate binding. First, calculations of water-accessible surface show a pocket at the copper site which is structurally complementary to D-galactose. For this model of the enzyme-substrate complex, O-6 of

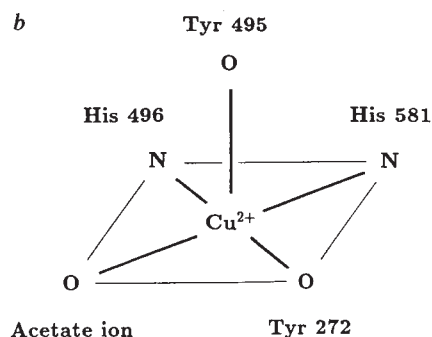
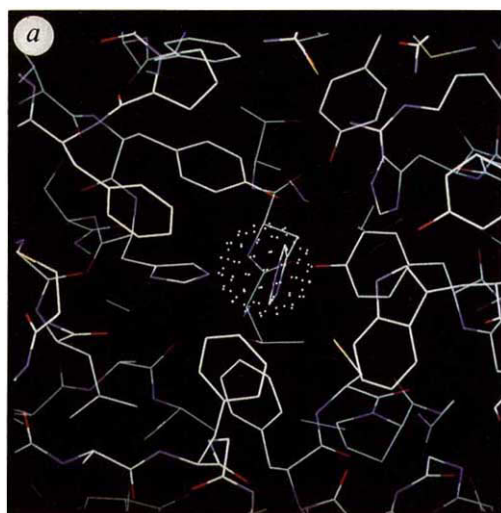
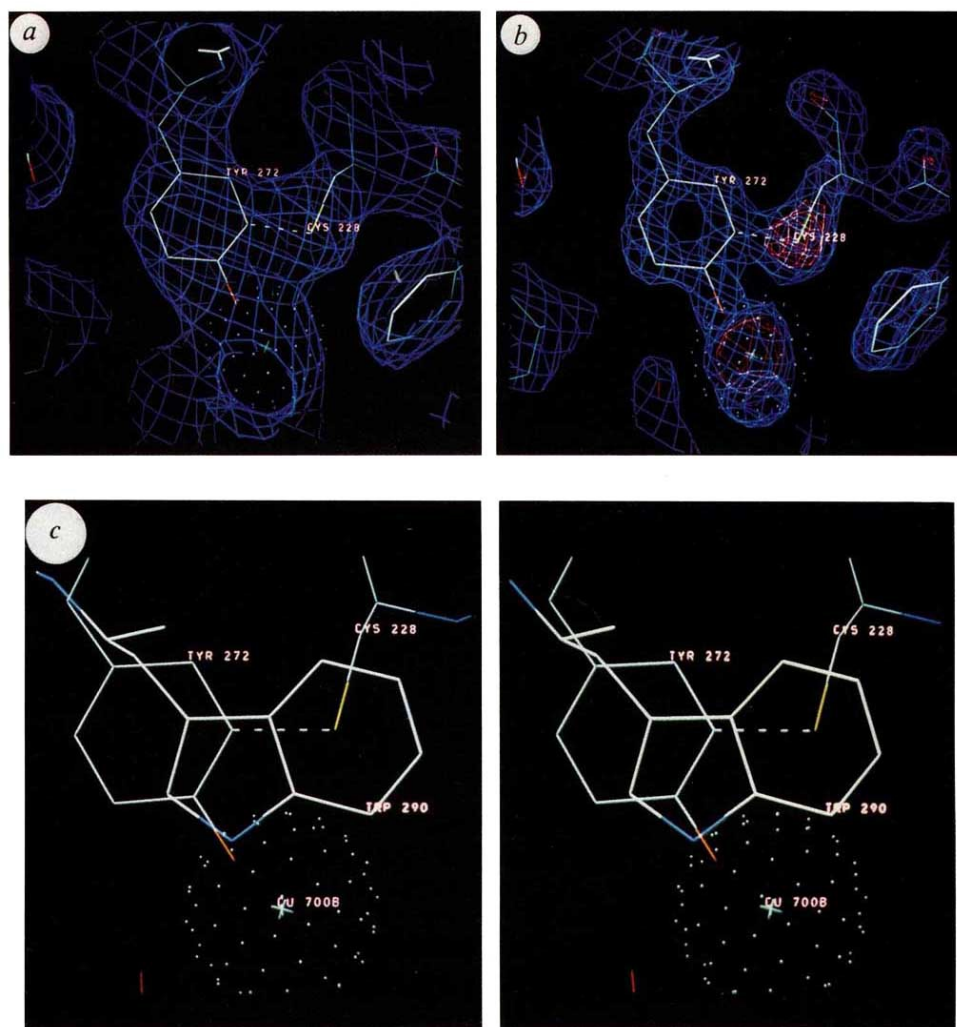


FIG. 2 *a*, The copper site of galactose oxidase, which is extremely rich in aromatic side chains. *b*, Schematic drawing of the copper coordination. At pH 4.5, Tyr 272, His 496, His 581 and an acetate ion form an almost perfect square. This square is distorted in PIPES buffer (pH 7.0), where the acetate ion is replaced by a water molecule 2.8 Å from the copper.

FIG. 3 *a*, The unusual thioether bond between Tyr 272 and Cys 228. The map shown is the 2.5 Å m.r. map after solvent flattening, which is not biased by the molecular model. *b*, The same view of the bond but with the refined 1.7 Å $2F_o - F_c$ map. The blue contour corresponds to 2σ level and the red to 5σ level. The refinement was carried out without assuming existence of the Cys-Tyr bond although the nonbonded contact restraints between the two side chains were removed. The bond length between C_ϵ and S_γ was refined to 1.84 Å, bond angles $C_\delta - C_\epsilon - S_\gamma$ to 119.4° , $C_\epsilon - C_\delta - S_\gamma$ to 120.6° and $C_\epsilon - S_\gamma - C_\beta$ to 105.1° . The *cis* conformation of the bond (the torsion angle $C_\delta - C_\epsilon - S_\gamma - C_\beta$ is 7°) results in the formation of a large planar group and suggests that it has partial double-bond character. Other side chain torsion angles are $\chi_1 = -83^\circ$ and $\chi_2 = -99^\circ$ (using C_ϵ of Tyr 272 as the fourth atom) for Cys 228, $\chi_1 = -68^\circ$ and $\chi_2 = -166^\circ$ for Tyr 272. *c*, Stereo view of the thioether bond and Trp 290 stacking over it.



the galactose could directly coordinate to the copper at the equatorial site occupied by acetate (Fig. 2*b*). This model could explain substrate specificity because O-4 of D-galactose could form a hydrogen bond to Arg 330 whereas, for D-glucose (not a substrate), steric hindrance would prevent binding. Second, substrate might bind to the free radical site rather than the copper. The stacking Trp 290 is more likely to be the binding site than Tyr 272 as it is exposed to solvent whereas Tyr 272 is buried.

The structure of the copper site in galactose oxidase is quite distinct from that reported from X-ray crystallographic analysis of other copper-containing oxidative enzymes, ascorbate oxidase¹² and superoxide dismutase¹³. Ascorbate oxidase catalyses the four-electron reduction of oxygen to water, and contains four coppers grouped as mononuclear and trinuclear species. The mononuclear site is structurally similar to that found in prokaryotic electron transfer proteins^{14,15} with two histidines, a cysteine and a methionine as ligands. The trinuclear site is composed of a copper-pair coordinated to a single copper having two histidines and a hydroxyl as its remaining ligands for tetra-coordination. It is this latter copper which appears to bear some structural similarity to the copper in galactose oxidase, and it has been speculated¹² that it forms part of the oxygen binding and reduction site. The Cu(II) site in oxidized superoxide dismutase has distorted square pyramidal symmetry with four 'in-plane' histidines and an axially coordinated water, but the reduced form has Cu(I) trigonally coordinated to three histidines¹⁶; these structures seem to be designed for binding the

substrate, O_2^- , to both the Cu(II) and Cu(I) forms of the enzyme. The structure of the Cu(I) site in substrate-reduced galactose oxidase is at present unknown, though inhibition studies with butylisocyanide indicate that Cu(I) in the reduced enzyme is the site of oxygen binding (D. M. Dooley & P.F.K., unpublished results). The active site of reduced galactose oxidase would thus allow oxygen binding at mononuclear copper, which has recently been shown to be chemically feasible¹⁷.

It is interesting to compare the radical site proposed for galactose oxidase with those in other proteins whose structures are known. In the H_2O_2 -oxidized intermediate of cytochrome *c* peroxidase, the haem site is rich in aromatic amino-acid side chains, as is the copper site of galactose oxidase. The radical site, Trp 191 (ref. 18), is 5.1 Å from the iron, and the side chain of Asp 235 interacts with both Trp 191 and the proximal haem ligand, His 175. Although Trp 191 is itself buried, several tyrosine residues are close to the surface, where cytochrome *c* is suggested to bind¹⁹. Electron relay from cytochrome *c* via 'surface' tyrosine(s) and Trp 191 to Fe^{4+} is clearly similar to the possible electron relay in galactose oxidase from substrate alcohols via Trp 290 and Tyr 272 to Cu^{2+} . In ribonucleotide reductase, the radical site has been identified as Tyr 122 (ref. 20), which is located 5.3 Å from the closest iron of a dinuclear iron centre and buried ~ 10 Å from the nearest surface²¹. The hydrophobic nature of this site perhaps together with delocalization of the unpaired species accounts for the remarkable stability of the radical. Again there are similarities to galactose oxidase.

We believe that galactose oxidase provides an ideal model for the detailed mechanistic and structural analysis of a protein free-radical in enzymes and has wide implications for our understanding of fundamental biological processes in which such free radicals have been functionally implicated²². These include DNA and prostaglandin biosynthesis^{21,23,24}, the photo-synthetic reaction centre redox reaction²⁵ and cytochrome *c* peroxidase²⁶. □

Received 25 September; accepted 13 December 1990.

1. Kosman, D. J. in *Copper Proteins and Copper Enzymes* Vol. 1 (ed. Lontie, R.) 1–26 (CRC Press, Boca Raton publishers, Boca Raton 1984).
2. Dooley, D. M. *Life Chemistry Reports* **5**, 91–154 (1984).
3. Hamilton, G. A. *et al. J. Am. chem. Soc.* **100**, 1899–1912 (1978).
4. van der Meer, R. A., Jongejans, J. A. & Duine, J. A. *J. biol. Chem.* **264**, 7792–7794 (1989).
5. Whittaker, M. M. & Whittaker, J. W. *J. biol. Chem.* **263**, 6074–6080 (1988).
6. Whittaker, M. M., DeVito, V. L., Asher, S. A. & Whittaker, J. W. *J. biol. Chem.* **264**, 7104–7106 (1989).
7. Kosman, D. J., Ettinger, M. J., Weiner, R. E. & Massaro, E. J. *Archs Biochem. Biophys.* **165**, 456–467 (1974).
8. Vellieux, F. M. D. *et al. EMBO J.* **8**, 2171–2178 (1989).
9. Varghese, J. M., Laver, W. G. & Colman, P. M. *Nature* **303**, 35–40 (1983).

10. Tresselt, P. & Kosman, D. J. *Biochem. biophys. Res. Commun.* **92**, 781–786 (1980).
11. Lerch, K. *J. biol. Chem.* **257**, 6414–6419 (1982).
12. Messerschmidt, A. *et al. J. molec. Biol.* **206**, 513–529 (1989).
13. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. *J. molec. Biol.* **160**, 181–217 (1982).
14. Guss, J. M. & Freeman, H. C. *J. molec. Biol.* **169**, 521–563 (1982).
15. Adman, E. T. & Jensen, L. H. *Israel J. Chem.* **21**, 8–12 (1981).
16. Blackburn, N. J. *et al. Biochem. J.* **219**, 985–990 (1984).
17. Jacobson, R. R. *et al. J. Am. Chem. Soc.* **110**, 3690–3692 (1988).
18. Sivaraja, M., Goodin, D. B., Smith, M. & Hoffman, B. M. *Science* **245**, 738–740 (1989).
19. Poulos, T. L. & Kraut, J. *J. biol. Chem.* **255**, 8199–8205 (1980).
20. Larsson, Å. & Sjöberg, B.-M. *EMBO J.* **5**, 2037–2040 (1986).
21. Nordlund, P., Sjöberg, B.-M. & Eklund, H. *Nature* **345**, 593–598 (1990).
22. Stubbe, J. A. *et al. Rev. Biochem.* **58**, 257–285 (1989).
23. Sjöberg, B.-M., Reichard, P., Gräslund, A. & Ehrenberg, A. *J. biol. Chem.* **253**, 6863–6865 (1978).
24. Karthien, R., Dietz, R., Nastainczyk, W. & Ruf, H. H. *Eur. J. Biochem.* **171**, 313–320 (1988).
25. Barry, B. A. & Babcock, G. T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7099–7103 (1987).
26. Prince, R. C. & George, G. M. *Trends biochem. Sci.* **15**, 170–172 (1990).
27. Kabsch, W. *J. appl. Crystallogr.* **21**, 916–924 (1988).
28. Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
29. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
30. Priestle, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).

ACKNOWLEDGEMENTS. We thank V. I. C. Blakeley for help in preparation of the enzyme, C. W. G. Boys for help in the early stages of the project, and D. R. Flower for help with the graphics programs. Computing and X-ray crystallographic facilities were provided through the SERC Molecular Recognition Centre at Leeds. The coordinates will be deposited in the Brookhaven Protein Data Bank.

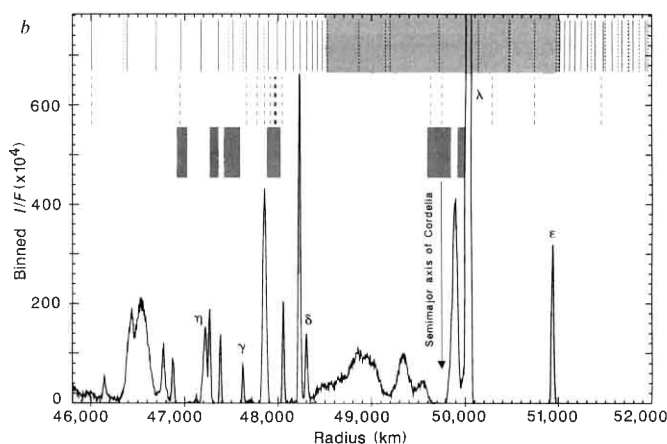
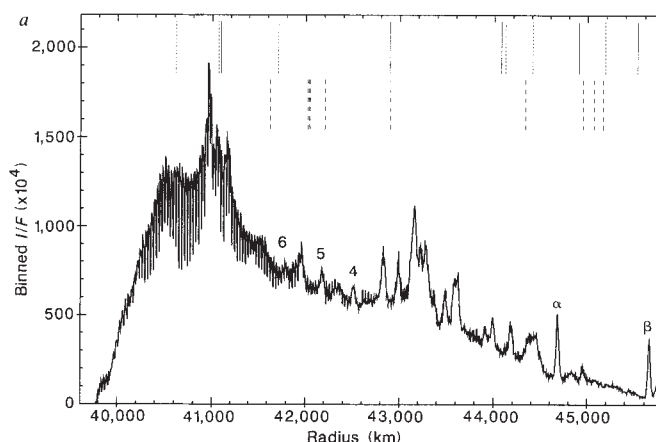
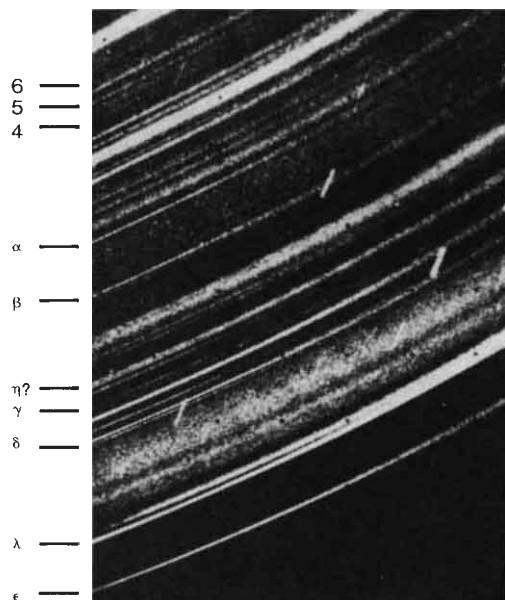
ERRATUM

Orbits of shepherd satellites deduced from the structure of the rings of Uranus

Carl D. Murray & Robert P. Thompson

Nature **348**, 499–502 (1990).

IN this Article in the 6 December 1990 issue, an error in the *Nature* office led to mislabelling of a ring in Fig. 1. A corrected version appears below and in reprints. In addition some labelling was omitted from Fig. 2, shown on the right.



Nucleic acid sequence-based amplification

J. Compton

Nucleic acid sequence-based amplification (NASBA) is a primer-dependent technology that can be used for the continuous amplification of nucleic acids in a single mixture at one temperature.

NUCLEIC acid amplification has rapidly become one of the most important tools for molecular biologists. Technology based on the polymerase chain reaction (PCR) uses repeated thermal denaturation of double-stranded DNA and extension by a thermostable DNA polymerase of two primers, each annealed to the complementary template, to achieve DNA amplification of greater than one million-fold in about three to four hours¹. This method requires rapid changes in sample temperature, for which specialized thermal cycling instruments are usually required.

Nucleic acid sequence-based amplification offers a simple and rapid alternative method for nucleic acid amplification, which can yield an RNA amplification of a billion-fold in about two hours (L. Malek, unpublished data). In addition, as NASBA is a continuous, isothermal process, specialized equipment is not required.

NASBA amplification process

A standard NASBA reaction mixture contains T7 RNA polymerase, RNase H, AMV (avian myeloblastosis virus) reverse transcriptase, nucleoside triphosphates, two specific primers and appropriate buffer components. Primer 1 is about 45 bases in length, with an average of 20 bases at the 3' end that are complementary to the 3' side of the target sequence. The 5' end of this primer contains a promoter sequence that is recognized by T7 RNA polymerase. Primer 2 is about 20 bases in length and is derived from the opposite (5' direction) side of the target sequence.

One of the properties of the NASBA system is the ability to amplify directly specific sequences of single-stranded RNA. When using a natural RNA template (see Fig. 1), amplification is achieved by adding the template to the standard NASBA reaction described above. In the so-called 'non-cyclic' phase of NASBA, primer 1 anneals to the RNA target sequence. AMV reverse transcriptase, which is present in the reaction mixture, uses the deoxynucleoside triphosphates, also present, to extend the 3' end of primer 1, thereby forming a cDNA copy of the template and resulting in an RNA:DNA hybrid. RNase H hydrolyses RNA from such RNA:DNA hybrids. Consequently, the original RNA is destroyed, leaving a single strand of DNA to which primer 2 anneals. Reverse transcriptase synthesizes the second DNA strand, rendering the promoter region

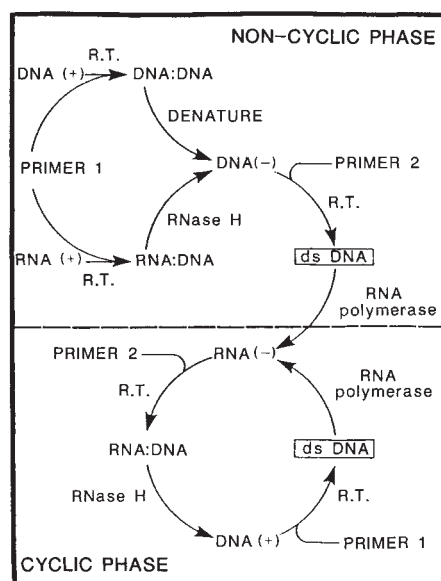


FIG. 1 Process for the continuous, homogeneous, isothermal amplification of RNA or DNA.

double-stranded.

The third enzyme in the mixture, T7 RNA polymerase, transcribes RNA copies from the now transcriptionally-active promoter, generating as many as 100 copies from each template molecule. Each new RNA molecule is now available as template for reverse transcriptase in the so-called 'cyclic' phase of the NASBA process (see Fig. 1). In the 'cyclic' phase, primer 2 binds to the template first and the action of reverse transcriptase extends it, thus generating an RNA:DNA hybrid as before. RNase H now hydrolyses the RNA strand, primer 1 binds to the resulting single-stranded DNA, and reverse transcriptase synthesizes DNA, again yielding a transcriptionally-active promoter. Although the previously mentioned steps have been detailed separately for clarity, all occur simultaneously enabling an amplification of a given target sequence to be achieved in a continuous, homogeneous and isothermal fashion (see Fig. 1).

Features of the technology

Given the apparent complexity of the interrelated enzymatic activities that are required for NASBA to function, the process might be expected to be intricate and/or temperamental. In practice, however, NASBA is surprisingly robust. Prepared template samples are added to NASBA reaction mixtures that contain kit components and ap-

propriate primers. Incubation at a single temperature for 1.5 to 2 hours is all that is required to achieve an amplification of about 10^9 -fold (L. Malek, unpublished data). In a model system, template input of 10^4 molecules results in the synthesis of a total of about 2×10^{13} full-length RNA molecules, or 1.3 μ g of RNA in two hours — an amount that can easily be seen on an ethidium bromide-stained agarose gel.

The NASBA process requires fewer 'cycles' than PCR to produce a desired amplification. When using PCR, the number of molecules doubles with each step, and so it requires approximately 20 cycles to produce an amplification of one million-fold. With NASBA, however, 10–100 copies of RNA are generated in each transcription step, so only four to five 'cycles' are required to achieve a similar amplification. Consequently, both the total incubation time and the overall error frequency are reduced with NASBA. Errors that are inherent in some enzymatic activities (for example, reverse transcriptase) are cumulative, and so fewer cycles should reduce the likelihood of these errors occurring. The fidelity of the NASBA process has been tested by sequencing the RNA product directly from a NASBA reaction. Of the sequence derived in this way, 90 per cent is readable; the 10 per cent of unreadable sequence appears to be due either to inherent enzyme 'stops' in the sequence or to strand degradation, which is to be expected with this type of non-purified sample.

The specificity of the NASBA process has been confirmed in model systems by the analysis of reactions performed in the presence and absence of carrier DNA. Using 10^4 molecules of template, the presence of 1 μ g of human placental DNA (equivalent to about 2×10^6 cells) does not significantly change the product yield after two hours of amplification. Using lower levels of input (for example, 10^2 molecules), however, significantly reduces the total yield of specific product in a two-hour reaction. Nevertheless, when the reaction products are compared to a known amount of standard on a slot-blot, the amplification efficiency is still in the order of 10^9 (L. Malek, unpublished data). 'No template' control reactions remain unchanged by the addition of excess heterologous DNA and do not produce characteristic product bands by either agarose gel or slot-blot analysis.

The sensitivity of the NASBA technology has been tested using several series of differ-

ent template dilutions. Consistently, samples with as few as ten molecules of input produce positive results, and samples with even fewer input molecules (as determined statistically) are detectable. Samples of HIV-infected cells or infectious HIV virus particles, diluted in human plasma, have also been detected with the same level of sensitivity as the model systems (P. Lens, unpublished data). False positives are not seen, and specific product bands are seen only in samples containing template.

Nucleic acid sequence-based amplification can be used for samples other than single-stranded RNA. Double-stranded nucleic acids can also be used as templates, but must first be converted to single-stranded molecules that are compatible with the process. Nucleic acids with discrete ends (for example, those generated by restriction enzyme digestions) are also compatible with the NASBA process and enter directly into the 'cyclic' phase, if they are first rendered single-stranded (see Fig. 1).

Applications

This technology has applications as a research tool, particularly in the area of RNA research. For example, NASBA simplifies both the detection and sequencing of mRNA sequences. With the absence of thermal denaturation in the NASBA process, contaminating genomic DNA is not amplified and, therefore, is not a concern. As NASBA does not require specialized equipment, it is likely to be suitable for researchers who have either limited requirements for nucleic acid amplification or restricted budgets.

In addition to its use as a research tool, NASBA also has clinical applications. Human *in vitro* diagnostic uses of NASBA include the screening of infectious diseases and genetic disorders, the diagnosis and monitoring of cancer, and testing for disease susceptibility. Data obtained using *in vitro* infected samples indicate that, in practice, both the specificity and sensitivity are consistent with that of model systems. As few as three HIV particles could be detected in a 100- μ l sample of reconstituted human plasma. An analysis of samples taken from patients with AIDS confirmed the sensitivity of the technology (P. Lens, unpublished data).

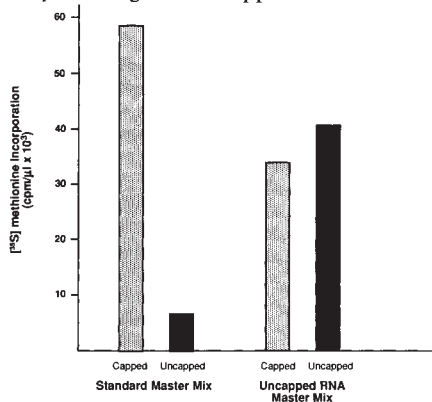
Nucleic acid sequence-based amplification technology can also be applied to food screening, veterinary diagnostics, aquaculture and, to some degree, to forensic science. □

Jean Compton is at Canguene Corporation, 3403 American Drive, Mississauga, Ontario L4V 1T4, Canada. For more information, fill in reader service number 100. NASBA is a trademark of Canguene Corporation with patents pending. The polymerase chain reaction is covered by US patents issued to Cetus Corporation.

At the cutting edge

A mouse mutagenesis assay system, a gel-digesting enzyme for DNA/RNA recovery and a handy electroporation device are a selection of this week's supplies and equipment for the molecular biologist.

ACCORDING to Ambion, its new Retic Lysate IVT Kit, a **rabbit reticulocyte cell-free system** for *in vitro* translation of protein from added mRNA, offers a significant improvement over existing nuclease-treated reticulocyte lysate translation systems (*Reader Service No. 101*). Two of the improvements are a new buffer system or "Master Mix", which allows the efficient translation of uncapped *in vitro* transcripts and a production procedure that results in significantly more active lysate, says Ambion. The \$95 (US) Retic Lysate IVT Kit is supplied with four different Master Mixes: two standard mixes (-Met and -Leu) and two uncapped RNA mixes (-Met and -Leu) for uncapped *in vitro* transcripts. These uncapped RNA Master Mixes allow the efficient translation of uncapped mRNAs. Uncapped globin transcripts, using the uncapped RNA Master



Comparison of translation of capped and uncapped globin transcripts using standard and uncapped RNA Master Mixes.

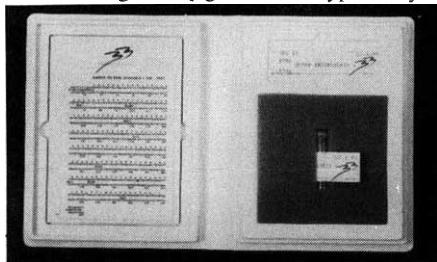
Mix, are translated at up to 70 per cent of the efficiency of capped globin using the standard Master Mix. Many reticulocyte lysates suffer from premature termination of initiation, which results in termination of protein synthesis after about an hour. Ambion's lysate continues to synthesize proteins for at least two hours, which, the company says, can be important in experiments where a large mass amount of translation product is advantageous or when a protein of high molecular weight is being synthesized. The lysate incorporates 48 per cent of added [³⁵S]methionine into protein using the control mRNA supplied with the kit.

SeaKem Gold agarose from FMC Bio-Products is intended for fast separations of either DNA fragments of greater than 1,000 bp by conventional electrophoresis or megabase DNA by pulsed-field gel electrophoresis (*Reader Service No. 102*). Although the increased speed of separation

that can be achieved with SeaKem Gold depends on both user technique and running conditions, the range of the increase is between 5 and 40 per cent, which is still faster than either standard or other 'fast-running' agarose products, says FMC. Gels that are easy to handle can be prepared from concentrations of agarose as low as 0.3 per cent. DNA that is recovered from SeaKem Gold agarose is compatible with restriction and ligation enzymes.

Brush up on your DNA sequencing technique with the three-part **sequencing support service** from United States Biochemical (*Reader Service No. 103*). This service consists of a DNA sequencing guide, reference material and video. The guide concentrates on how to obtain optimal sequencing results with various templates when using the chain-termination method. Articles on the enzymology of DNA sequencing (theory) and DNA sequencing methodologies are included in the literature collection. The one-hour video presentation explores both the theory and the technique of basic and advanced chain-termination (dideoxy) DNA sequencing. The literature collection and sequencing guide will be updated periodically.

Over 60 **designer genes**, coding for cytokines and growth factors (human and murine), endothelial adhesion molecules and viral components, are offered by British Bio-technology (*Reader Service No. 104*). The range includes interleukins 1-9, TGF β , CSFs, PDGF, ICAM-1, ELAM-1, PECAM and HIV regulatory genes. Two types of syn-



Designer genes by mail order.

thetic genes are available: fully designed genes, which are suitable for expression in *Escherichia coli*, and natural genes for expression in eukaryotic systems. Intended applications include expression studies, hybridization probes, structure-function studies and the construction of multi-functional chimaeras.

Arrow Scientific is the UK distributor for the range of reagents for **DNA and RNA synthesis** that have been developed by

1. Saiki, R.K. et al. *Science* **239**, 487-491 (1988).

ADVERTISEMENTS

American Bionetics (*Reader Service No. 105*). Over 250 products are offered. CED phosphoramidites, which are prepared with a β -cyanoethyl phosphate protecting group in place of the standard methyl group, can be substituted directly for standard phosphoramidites without alteration to cycle programmes used on DNA synthesizers. They can be fully deprotected in one step using ammonium hydroxide, thereby eliminating the need for treatment with the toxic and corrosive reagent thiophenol at the completion of DNA synthesis, says the company. Deoxyinosine CED phosphoramidite, which is designed for the convenient synthesis of probes containing degeneracies, is prepared from 2'-deoxyinosine by a process involving a β -cyanoethyl phosphate protecting group; this conveys significant performance advantages over the previously used *O*-methyl group, says the company. A selection of reagents for RNA synthesis is also available, including four RNA nucleoside phosphoramidite monomers and four protected ribonucleoside derivatized solid supports, which are available in both 500 and 1,000 Å pore sizes.

Detecting mutations

Using transgenic mouse technology, Stratagene has developed a short-term *in vivo* **mutagenesis assay system** called Big Blue, which is designed to assess the effects of chemical exposure on germ-line and somatic cells (*Reader Service No. 106*). Big Blue uses a bacteriophage lambda shuttle vector that has been integrated, via microinjection, into the genome of inbred mice. The vector contains the *LacI^q* gene, which serves as a target for mutagenesis. After administration of the test compound to the mouse, the shuttle vector containing the target gene can be recovered from genomic DNA isolated from any tissue of the mouse. This target DNA is then packaged into empty phage heads. Mutations within the *LacI^q* gene can be identified by infecting host *E. coli* with the packaged phage and plating on indicator agar containing a chromogenic substrate that is specific for the *Lac* gene system. The mutant frequency of the test compound is determined by a colour reaction. The Big Blue system offers several advantages, says Stratagene. The effects of chemical exposure on a whole animal, not just selected cells in an *in vitro* system, can be determined. The assay is fast. After administering the dose regimen, the assay can be performed in about three days. And, the target vector has functions that permit the *in vivo* excision of the *LacI^q* target gene into a plasmid; this allows mutation characterization by means of sequence analysis.

Clontech has introduced new Mouse *Ki-ras* Onco-Lyzer Kits and Amplimer Sets for the polymerase chain reaction **analysis of mouse *Ki-ras* oncogenes** (*Reader Service No. 107*). The \$395 (US) Onco-Lyzer Kits and Amplimer Sets are available to detect

mutations of *Ki-ras* oncogenes in codons 12 (exon 1), 61 (exon 2) and 116-119 (exon 3) in a mouse model system. Each kit contains 5 π and 3 π amplimers, AmpliTaq DNA polymerase, PCR buffer, dNTPs and control template. Onco-Lyzer Kits and Amplimer Sets for PCR analysis of Human *Ki-ras*, *H-ras*, *N-ras*, *c-myc* and *HER-2/neu* oncogenes are already available.

Assorted kits

International Biotechnologies introduces the FLAG Kit for the **expression, detection and purification of recombinant proteins** (*Reader Service No. 108*). FLAG technology involves the fusion of a hydrophilic, eight amino acid FLAG peptide to a cloned protein that is expressed by the pFLAG-1 vector. Using an anti-FLAG monoclonal antibody coupled to agarose, the FLAG fusion protein is affinity purified. After purification, the native protein is recovered with greater than 99 per cent efficiency by removal of the FLAG peptide with enterokinase, says IBI. The protein is recovered in a biologically active form due to the small size of the FLAG peptide.

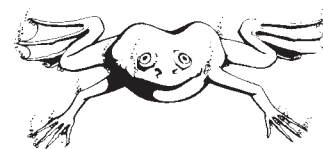
A **footprinting system for transcription factor AP1** is now available from Promega, which allows the user to determine whether a gene of interest contains a *c-jun* binding site (*Reader Service No. 109*). The footprinting system is supplied with purified bacterially-expressed *c-jun* protein, a test plasmid and all the reagents that are necessary to perform footprinting experiments. The test plasmid is cut with a restriction enzyme and is dephosphorylated using calf intestinal alkaline phosphatase. The DNA is then ³²P-labelled with T4 polynucleotide kinase and the label is then removed from one end by treatment with a second restriction enzyme. This results in a singly end-labelled probe. Following binding of *c-jun* protein, the probe is partially digested with DNase I and the resulting fragments are analysed on a sequencing gel. The region of the probe that is protected from DNase by binding of the *c-jun* protein is revealed by a region of missing bands in the degradation profile. Footprinting systems for SP1, AP2 and TFIID are also available.

Sanbio of The Netherlands has developed the NucleoScan kit for the **measurement of small quantities of double-stranded DNA in cells and tissues** (*Reader Service No. 110*). The kit eliminates the need for complex extraction procedures. In the two-step procedure, the DNA is liberated and then stained with a fluorescent label. The test takes about two hours, says Sanbio, and can be used to measure quantities of dsDNA between 0.5 µg and 30 µg.

Enzyme ensemble

GELase is a new **gel-digesting enzyme** preparation from Epicentre Technologies for the quantitative recovery of DNA and

SINGER
INSTRUMENTS



COMPLETE SYSTEMS FOR
XENOPUS
OOCYTE INJECTION

INT.SALES: +44.984.40226
FAX: +44.984.41166

Singer Instrument Co. Ltd.
Roadwater, Watchet, Somerset,
England TA23 0QL. Tel: Washford
(0984) 40226. Telex: 337492 Comcab G

ACCESS, VISA and
AMERICAN EXPRESS

Reader Service No.50

TIB MOLBIOL BERLIN
CUSTOM DNA SYNTHESIS

DM per base
5,-

DM 95,- per oligo

FAX (0049) 30 6261960

Reader Service No.79

METABOLITES analysed
rapidly, enzymatically, using micro-samples

with
the



ANALOX GM7 MICRO-STAT

A wide range of biological fluids including whole blood are acceptable for direct injection. Turbidity no problem. Sample size 2.5 - 25 µl. Results in 20 seconds.

glucose	lactate	urea
alcohol	pyruvate	creatinine
cholesterol	ammonia	3-OH butyrate
uric acid		acetoacetate

Analox Instruments Ltd

8 Goldhawk Industrial Estate
Brackenhurst Road Tel: 081-749 7635
Hammersmith, London W6 0BQ Fax: 081-740 6608

Reader Service No.16

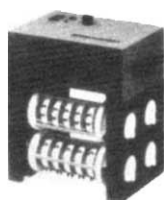
Copies of articles from
this publication are now
available from the UMI
Article Clearinghouse.

UMI Article
Clearinghouse

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

ADVERTISEMENTS

DIANORM
EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

DIANORM
Genate
P.O. Box 650726
D-8000 Munich 65,
FRG.

Reader Service No.49

Custom DNA

Purified and Shipped in 48 hours,
Worldwide.

100 ug, \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393
In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.32

Synthetic Oligonucleotides

Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days

£4.00 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, aminoalkylation,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56

Professional Custom Services

DNA SYNTHESIS

PROTEIN SEQUENCING

PEPTIDE SYNTHESIS

Rapid Service · Guaranteed Quality Ample
Quantity

Competitive Prices · Free Quotations

Worldwide Delivery

V.G. Services,

1200 Wonderland Rd,

S. Bldg. #7, Unit #2, London,

Ontario Canada N6L 1A8

Call/Fax: (519) 652-3758 Alt. Fax: (519) 652-9622

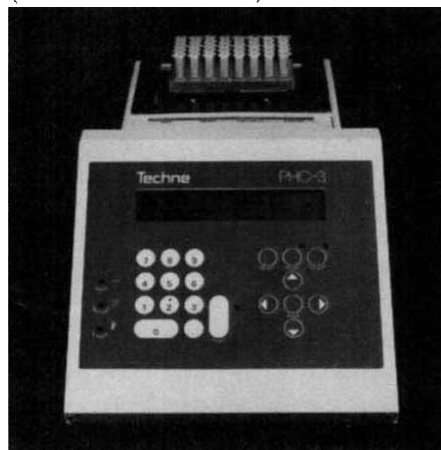
RNA from low-melting point agarose gels (*Reader Service No. 111*). GELase digests LMP-agarose gels in Tris-acetate or Tris-borate buffers, yielding a clear liquid that does not become viscous or gel, even on cooling, says Epicentre. When using the recommended quantity of GELase, the incubation time required to digest a LMP-gel slice is about one hour. The resulting solution of DNA or RNA is then ready for further manipulation. Alternatively, as the gel digestion products are soluble in alcohol, the DNA/RNA can be recovered by ethanol precipitation. The gentleness of the method for recovering nucleic acids from agarose gels permits the isolation of high molecular weight DNA for genome mapping or sequencing, cosmid cloning and other uses, as well as low molecular weight DNA or RNA.

For DNA sequencing at elevated temperatures, Bio-Rad recommends its **DNA sequencing enzyme** *Bst* polymerase (*Reader Service No. 112*). *Bst* polymerase is based on DNA polymerase I, large fragment, purified from the thermophilic bacterium *Bacillus stearothermophilus*. The enzyme can be used with either ³²P or ³⁵S label; the high signal allows short exposure times of 8 hours for ³²P, overnight for ³⁵S, says Bio-Rad. Standard isotopic fluorescent protocols can be used. Sequencing can be performed with M13 vectors or double-stranded plasmids. *Bst* polymerase is compatible with the nucleotide analogue 7-deaza dGTP.

Twenty two new **restriction enzymes** are offered by International Laboratory Services (*Reader Service No. 113*). The line up includes enzymes with specificities for CGPuPy/CG (*BsiE I*) and CCNNNNN/NGG (*BsiYI*). Of the 22 enzymes, several are thermophilic and are suitable for cutting genomic DNA with high degrees of secondary structure, says ILS. All enzymes are supplied in 100-, 200- or 1,000-unit quantities.

Laboratory hardware

To complement Techne's existing line of **thermal cyclers**, the company has introduced the **PHC-3 Programmable Dri-Block cycler** (*Reader Service No. 114*). The PHC-3 con-



Thermal cycling with the PHC-3.

tains Peltier thermo-electric modules, which provide heating and cooling over a temperature range of 4° C to 99° C, with a block uniformity of 1°C, says Techne. PHC-3 has 99 user-selectable programmes, each of which provides up to 99 segments (temperature, hold time and ramp rate). Programmes can also be linked together. The user is provided with a choice of interchangeable blocks that can hold 40 × 0.5-ml microcentrifuge tubes, Techne Hi-Temp 96 or Falcon round-bottom multiwell plates, and a flat plate for microscope slides for *in vitro* experiments.

Electroporation involves exposing a cell to a brief high-voltage pulse, thereby temporarily rendering the cell membrane permeable to DNA and other macromolecules. The **Porator electroporation device** from Invitrogen is designed to provide efficient and reliable electroporation of mammalian and bacterial cells (*Reader Service No. 115*). Invitrogen says the Porator allows researchers to achieve high levels of transient expression and stable transformation of mammalian cells, and produce competent *E. coli* cells with higher transformation efficiencies than cells modified by chemical means. The instrument is supplied with five vials of electro-competent cells, ten 0.1-cm gap cuvettes and an instruction manual.

Faster data acquisition and data analysis of gels, films and membranes are features of the new **300B computing densitometer**



300B: for gel, film and blot analysis.

from Molecular Dynamics (*Reader Service No. 116*). This has been made possible by the installation of a powerful 16 MHz Intel 80386SX microprocessor to drive the optical scanning system and automatically store complete images of the scan area on the integral hard disk. The 300B uses a helium-neon laser and a light collection cylinder to provide readings over a wide range of optical densities. The instrument is designed to scan electrophoresis gels, autoradiographic films, and blotting membranes as large as 36 × 42 cm in less than three minutes; small samples are scanned more rapidly. As the 300B scans entire samples rather than single lines, it is useful for the quantitation of faint bands on 1- and 2-D gels. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.